ATSC Standard: Video – HEVC

Doc. A/341:2022-03
31 March 2022

Advanced Television Systems Committee
1300 I Street, N.W., Suite 400E
Washington, D.C. 20005
202-872-9160

The Advanced Television Systems Committee, Inc. is an international, non-profit organization developing voluntary standards and recommended practices for broadcast television and multimedia data distribution. ATSC member organizations represent the broadcast, professional equipment, motion picture, consumer electronics, computer, cable, satellite, and semiconductor industries. ATSC also develops implementation strategies and supports educational activities on ATSC standards. ATSC was formed in 1983 by the member organizations of the Joint Committee on Inter-society Coordination (JCIC): the Consumer Technology Association (CTA), the Institute of Electrical and Electronics Engineers (IEEE), the National Association of Broadcasters (NAB), the Internet & Television Association (NCTA), and the Society of Motion Picture and Television Engineers (SMPTE). For more information visit www.atsc.org.

Note: The user's attention is called to the possibility that compliance with this standard may require use of an invention covered by patent rights. By publication of this standard, no position is taken with respect to the validity of this claim or of any patent rights in connection therewith. One or more patent holders have, however, filed a statement regarding the terms on which such patent holder(s) may be willing to grant a license under these rights to individuals or entities desiring to obtain such a license. Details may be obtained from the ATSC Secretary and the patent holder.

Implementers with feedback, comments, or potential bug reports relating to this document may contact ATSC at https://www.atsc.org/feedback/.

Revision History

Version	Date
Candidate Standard approved	23 December 2015
A/341:2017 Standard approved	19 May 2017
Amendment No. 1 approved	27 December 2017
Amendment No. 2 approved	28 December 2017
Amendment No. 3 approved	29 December 2017
Amendment No. 4 approved	24 January 2017
A/341:2018 (a rollup of previous Amendments to A/341:2017)	24 January 2018
Amendment No. 1 approved	9 March 2018
Amendment No. 2 approved	12 March 2018
Corrigenda No. 1 approved	28 August 2018
A/341:2019 Standard approved	14 February 2019
Amendment No. 1 approved	2 March 2021
Amendment No. 2 approved	17 September 2021
A/341:2021 (a rollup of previous Amendments to A/341:2019)	17 September 2021
A/341:2022-03 published (references to ATSC documents updated)	31 March 2022

Table of Contents

1. SCOPE 1
   1.1 Introduction and Background 1
   1.2 Organization 1
2. REFERENCES 1
   2.1 Normative References 1
   2.2 Informative References 2
3. DEFINITION OF TERMS 3
   3.1 Compliance Notation 4
   3.2 Treatment of Syntactic Elements 4
      3.2.1 Reserved Elements 4
   3.3 Acronyms and Abbreviations 4
   3.4 Terms 5
   3.5 Extensibility 6
      3.5.1 Backward-compatible Extensibility Mechanisms 6
      3.5.2 Non-backward-compatible Extensibility Mechanisms 6
      3.5.3 Extensions with unknown compatibility 6
      3.5.4 Descriptor Processing Considerations 6
4. SYSTEM CAPABILITIES (INFORMATIVE) 6
5. ACTIVE FORMAT DESCRIPTION 6
6. SOURCE CODING SPECIFICATION 7
   6.1 HEVC (ISO/IEC 23008-2) 7
   6.2 Video Formats 7
      6.2.1 Legacy SD Video 7
      6.2.2 Interlaced HD Video 8
      6.2.3 Progressive Video 9
   6.3 Specific Constraints Regarding Additional Features 10
      6.3.1 Specific Constraints Regarding Spatial Scalable Coding 10
      6.3.2 Specific Constraints Regarding Video Transfer Characteristics 12
      6.3.3 Specific Constraints Regarding 3D 16
      6.3.4 Specific Constraints Regarding High Frame Rate Temporal Sub-Layering 17
   6.4 Bit Stream Specifications in Addition to Video Coding 19
      6.4.1 Carriage of Active Format Description (AFD) and Bar Data 19
      6.4.2 AFD and Bar Data 19
      6.4.3 Optional Carriage of Closed Captions in the Video Stream 19
      6.4.4 Carriage of Picture Timing Information with Coded Pictures 19
7. SIGNALING OF VIDEO CHARACTERISTICS 20
ANNEX A : VIDEO INPUTS (INFORMATIVE) 22
   A.1 Television production standards 22
ANNEX B : EXAMPLE PROGRESSIVE VIDEO FORMATS (INFORMATIVE) 23
   B.1 Example Progressive Video Formats 23
ANNEX C : FIXED/MOBILE HYBRID UHD 3D TELEVISION SYSTEM (INFORMATIVE) 24
   C.1 Overview 24

ANNEX D : MULTIPLE FRAME RATE TEMPORAL FILTERING (INFORMATIVE).....26
D.1 Multiple Frame Rate Temporal Filtering 26
D.1.1 Creation of a Multiple Frame Rate Temporal Filtering Stream 26
D.1.2 High Frame Rate Video Recovery 27
ANNEX E : METADATA BASED ON SMPTE ST 2094-10_DATA 29
E.1 Metadata Based on ST 2094-10_Data (Normative) 29
ANNEX F METADATA BASED ON SMPTE ST 2094-40 35
F.1 Theory of Operation (Informative) 35
F.1.1 Metadata 35
F.1.2 Basis OOTF 35
F.1.3 Reference Method for Receiver-side Tone Mapping using ST 2094-40 Metadata 36
F.1.4 Statistical Characteristics 44

Index of Figures and Tables

Figure C.1.1 Example configuration of a mobile HD right view and a fixed UHD left view. 24
Figure C.1.2 Example configuration of a mobile HD right view and a fixed UHD left view using SHVC. 25
Figure D.1.1 Creation of a Multiple Frame Rate Temporal Filtering stream. 27
Figure D.1.2 Recovery of the HFR video stream. 28
Figure F.1.1 Example OOTF. 36
Figure F.1.2 Tone-mapping system of an ST 2094-40-based device. 37
Figure F.1.3 Example of guided knee point mixing parameter function when $D \leq T$. 39
Figure F.1.4 The guided knee point when $D \leq T$. 39
Figure F.1.5 Example of guided knee point mixing parameter function when $T \leq D$. 40
Figure F.1.6 The guided knee points when $T \leq D$. 40
Figure F.1.7 Example of Bezier curve vector mixing coefficient function for $D \leq T$. 42
Figure F.1.8 The Bezier curve anchor control determining function for $T \leq D$. 42
Figure F.1.9 The guided Bezier curve when $NORM \leq D$. 43
Figure F.1.10 Examples of the continuity of different guided OOTF curves at the knee point. 43

Table 5.1 SMPTE ST 2016-1 Table 2 – Video Format Information 7
Table 6.1 Legacy SD Video Formats 8
Table 6.2 Interlaced HD Video Formats 9
Table 6.3 user_data_type_code 15
Table 6.4 Multiview View Position SEI Message 17
Table 6.5 Multiview Scene Information SEI Message 17
Table 7.1 Video Characteristics 21
Table A.1.1 Standard Video Production Formats 22
Table B.1.1 Example Progressive Video Formats 23
Table E.1.1 ST2094-10_data() 29
Table E.1.2 ext_dm_data_block() 29
Table E.1.3 ext_dm_data_block_payload() 30
Table E.1.4 Definition of Extended Display Mapping Metadata Block Type 31

ATSC Standard: Video – HEVC

1. SCOPE

This Standard describes the video coding constraints on ITU-T Rec. H.265 | International Standard ISO/IEC 23008-2 [2] ("HEVC") when it is used for video compression in the ATSC 3.0 Digital Television System.

Any other video coding technologies in the ATSC 3.0 system are documented in their own ATSC Standard(s). A/300 [23] provides references to the various video coding technology document(s) in the ATSC 3.0 system. Signaling of the video compression technology in use is defined in [15].

1.1 Introduction and Background

This Standard specifies HEVC-coded video when it is used in the ATSC 3.0 Digital Television System. It specifies the allowable emission formats as well as features such as Spatial Scalable Coding, High Dynamic Range, Wide Color Gamut, 3D, and temporal sub-layering.

Formats include Legacy SD video (Section 6.2.1) and Interlaced HD video (Section 6.2.2) for efficient support of existing content, as well as Progressive Video (Section 6.2.3) that allows the full range of available features.

AFD and Bar Data are defined such that the active area of the picture does not necessarily need to fill the entire coded area. (See Sections 5 and 6.4.)

1.2 Organization

This document is organized as follows:

- Section 1 – Outlines the scope of this document and provides a general introduction.
- Section 2 – Lists references and applicable documents.
- Section 3 – Provides a definition of terms, acronyms, and abbreviations for this document.
- Section 4 – Describes an overview of the system.
- Section 5 – Describes the video processing before video compression.
- Section 6 – Specifies source coding.
- Section 7– Specifies signaling.
- Annex A – Describes video input standards.
- Annex B – Provides example progressive video formats.
- Annex C – Provides guidance related to fixed/mobile hybrid 3DTV

2. REFERENCES

All referenced documents are subject to revision. Users of this Standard are cautioned that newer editions might or might not be compatible.

2.1 Normative References

The following documents, in whole or in part, as referenced in this document, contain specific provisions that are to be followed strictly in order to implement a provision of this Standard.

[1] IEEE: "Use of the International Systems of Units (SI): The Modern Metric System," Doc. SI 10, Institute of Electrical and Electronics Engineers, New York, NY.

[2] ISO/IEC: "Information technology – High efficiency coding and media delivery in heterogeneous environments – Part 2: High efficiency video coding," Doc. ISO/IEC 23008-2:2020[1].

[3] ITU: ITU-R Recommendation BT.601-6 (2007), "Encoding Parameters of Digital Television for Studios," International Telecommunications Union, Geneva.

[4] ITU: ITU-R Recommendation BT.709-5 (2002), "Parameter values for the HDTV standards for production and international programme exchange," International Telecommunications Union, Geneva.

[5] ITU: ITU-R Recommendation BT.2020-1 (2014), "Parameter values for ultra-high definition television systems for production and international programme exchange," International Telecommunications Union, Geneva.

[6] SMPTE: "Format for Active Format Description and Bar Data," Doc. SMPTE ST 2016-1 (2009), Society of Motion Picture and Television Engineers, White Plains, NY.

[7] SCTE: "AVC Video Constraints for Cable Television, Part 1 – Coding," Doc. ANSI/SCTE 128-1 2013, Society of Cable Telecommunications Engineers, Exton, PA.

[8] CTA: "Digital Television (DTV) Closed Captioning," Doc. ANSI/CTA-708-E, Consumer Technology Association, Arlington, VA, August 2013.

[9] SMPTE: "Electro-Optical Transfer Function for High Dynamic Range Reference Display" Doc: ST 2084 (2014), Society of Motion Picture and Television Engineers, White Plains, NY.

[10] SMPTE: "Dynamic Metadata for Color Volume Transformation – Application #1," Doc. ST 2094-10 (2016), Society of Motion Picture and Television Engineers, White Plains, NY.

[11] ETSI: "High-Performance Single Layer High Dynamic Range (HDR) System for use in Consumer Electronics devices; Part 1: Directly Standard Dynamic Range (SDR) Compatible HDR System (SL-HDR1)", Doc. ETSI TS 103 433-1 V1.2.1 (2017-08), European Telecommunications Standards Institute, Sophia Antipolis Cedex, France.

[12] SCTE: "HEVC Video Constraints for Cable Television Part 1-Coding" Doc. ANSI/SCTE 215-1 2015, Society of Cable Telecommunications Engineers, Exton, PA.

[13] SMPTE: "Dynamic Metadata for Color Volume Transformation – Application #4," Doc. ST 2094-40:2020, Society of Motion Picture and Television Engineers, White Plains, NY.

[14] CTA: "A DTV Profile for Uncompressed High Speed Digital Interfaces," Doc. CTA-861-H (December 2020), Consumer Technology Association, Arlington, VA.

2.2 Informative References

The following documents contain information that may be helpful in applying this Standard.

[15] ATSC: "ATSC Standard: Signaling, Delivery, Synchronization, and Error Protection," Doc. A/331:2022-03, Advanced Television Systems Committee, Washington, DC, 31 March 2022.

[16] SMPTE: "Ultra High Definition Television — Image Parameter Values for Program Production," Doc. SMPTE ST 2036-1 (2013), Society of Motion Picture and Television Engineers, White Plains, NY.

[1] Also published by ITU as ITU-T Recommendation H.265.

[17] SMPTE: "Standard for Television—1920 x 1080 Scanning and Analog and Parallel Digital Interfaces for Multiple Picture Rates," Doc. SMPTE ST 274 (2008), Society of Motion Picture and Television Engineers, White Plains, NY.

[18] SMPTE: "Standard for Television—1280 x 720 Progressive Image Sample Structure, Analog and Digital Representation and Analog Interface," Doc. SMPTE ST 296 (2012), Society of Motion Picture and Television Engineers, White Plains, NY.

[19] SMPTE: "SDTV Component Video Signal Coding 4:4:4 and 4:2:2, for 13.5 MHz and 18 MHz Systems," Doc. SMPTE ST 125 (2013), Society of Motion Picture and Television Engineers, White Plains, NY.

[20] ETSI: "Digital Video Broadcasting (DVB); Specification for the use of Video and Audio Coding in Broadcasting Applications based on the MPEG-2 Transport Stream," Doc. ETSI TS 101 154 V2.1.1 (2015-03), European Telecommunications Standards Institute, Sophia Antipolis Cedex – France.

[21] CTA: "Active Format Description (AFD) and Bar Data Recommended Practice," Doc. CTA-CEB16-A, Consumer Technology Association, Arlington, VA, July 2012.

[22] DTG: "Digital Receiver Implementation Guidelines and Recommended Receiver Reaction to Aspect Ratio Signaling in Digital Video Broadcasting, v2.0.0", January 2012. Digital TV Group, London, England.

[23] W3C: "TTML Text and Image Profiles for Internet Media Subtitles and Captions (IMSC1)", [Candidate] Recommendation, W3C, www.w3.org.

[24] ATSC: "ATSC Standard: ATSC 3.0 System," Doc. A/300:2021, Advanced Television Systems Committee, Washington, DC, 7 July 2001.

[25] ATSC: "ATSC Standard: Captions and Subtitles," Doc. A/343:2022-03, Advanced Television Systems Committee, Washington, D.C., 31 March 2022.

[26] SMPTE: "Mastering Display Color Volume Metadata for High Luminance and Wide Color Gamut Images," Doc. ST 2086 (2014), Society of Motion Picture and Television Engineers, White Plains, NY.

[27] CTA: "HDR Static Metadata Extensions", Doc. CTA-861-G (November 2016), Consumer Technology Association, Arlington, VA.

[28] ITU: "Image parameter values for high dynamic range television systems for use in production and international programme exchange," Doc. Recommendation ITU-R BT.2100-0 (07/2016), International Telecommunications Union, Geneva.

[29] ITU: Report ITU-R BT.2390, "High dynamic range television for production and international programme exchange," International Telecommunications Union, Geneva.

[30] SMPTE: "Dynamic Metadata for Color Volume Transformation – Core Components," Doc. ST 2094-1 (2016), Society of Motion Picture and Television Engineers, White Plains, NY.

[31] SCTE: "HEVC Video Constraints for Cable Television Part 1-1 HDR," Doc. SCTE 215-1-1 2020b, Society of Cable Telecommunications Engineers, Inc., Exton, PA.

3. DEFINITION OF TERMS

With respect to definition of terms, abbreviations, and units, the practice of the Institute of Electrical and Electronics Engineers (IEEE) as outlined in the Institute's published standards [1] shall be used. Where an abbreviation is not covered by IEEE practice or industry practice differs from IEEE practice, the abbreviation in question will be described in Section 3.3 of this document.

3.1 Compliance Notation

This section defines compliance terms for use by this document:

shall – This word indicates specific provisions that are to be followed strictly (no deviation is permitted).

shall not – This phrase indicates specific provisions that are absolutely prohibited.

should – This word indicates that a certain course of action is preferred but not necessarily required.

should not – This phrase means a certain possibility or course of action is undesirable but not prohibited.

3.2 Treatment of Syntactic Elements

This document contains symbolic references to syntactic elements used in the video subsystem. These references are typographically distinguished by the use of a different font (e.g., restricted), may contain the underscore character (e.g., sequence_end_code) and may consist of character strings that are not English words (e.g., dynrng).

In addition, square brackets within syntactic elements (e.g., fixed_pic_rate_general_flag[i]) have a different meaning than square brackets shown in normal text (e.g., the HEVC video compression standard [2]). In the syntactic elements case, square brackets that enclose an index denote an element of a list, matrix, or array. Square brackets in normal text that include a number indicate a reference cited in Section 2 of this document.

3.2.1 Reserved Elements

One or more reserved bits, symbols, fields, or ranges of values (i.e., elements) may be present in this document. These are used primarily to enable adding new values to a syntactical structure without altering its syntax or causing a problem with backwards compatibility, but they also can be used for other reasons.

The ATSC default value for reserved bits is '1.' There is no default value for other reserved elements. Use of reserved elements except as defined in ATSC Standards or by an industry standards-setting body is not permitted. See individual element semantics for mandatory settings and any additional use constraints. As currently-reserved elements may be assigned values and meanings in future versions of this Standard, receiving devices built to this version are expected to ignore all values appearing in currently-reserved elements to avoid possible future failure to function as intended.

3.3 Acronyms and Abbreviations

The following acronyms and abbreviations are used within this document.

2D – Two dimensional

3D – Three dimensional

AFD – Active Format Description

ATSC – Advanced Television Systems Committee

CVS – Coded Video Sequence

fps – frames (or pictures) per second

HD – High Definition

HDR – High Dynamic Range

HEVC – High Efficiency Video Coding

HFR – High Frame Rate
HLG – Hybrid Log-Gamma
IP – Internet Protocol
NAL – Network Abstraction Layer
OFDM – Orthogonal Frequency-Division Multiplexing
OSD – On-Screen Display
PHY – Physical Layer
PLP – Physical Layer Pipe
PQ – Perceptual Quantization
SCG – Standard Color Gamut
SD – Standard Definition
SDR – Standard Dynamic Range
SFR – Standard Frame Rate
SEI – Supplemental Enhancement Information
SL-HDR1 – Single Layer High Dynamic Range part 1
SHVC – Scalable High Efficiency Video Coding
SPS – Sequence Parameter Set
UDP – User Datagram Protocol
UHD – Ultra High Definition
UHDTV – Ultra High Definition Television
VPS – Video Parameter Set
WCG – Wide Color Gamut

3.4 Terms

The following terms are used within this document.

active area – The portion of the video picture area that is being utilized for program content. Also referred to in CTA-CEB16 [21] as the useful image inside the video frame. Active area excludes letterbox bars and pillarbox bars.

additional view – Stereoscopic 3D video component using equivalent or lower resolution compared to the reference view video. The view can be provided together with the reference view to create an asymmetric 3D video. (e.g., reference view: UHD resolution, additional view: HD resolution).

High Dynamic Range – a feature that allows representation of video levels with much higher luminance values than is possible with traditional video methods. (See Section 6.3.2)

High Frame Rate – A picture frequency higher than 60 pictures per second.

Interlaced HD – A specific set of high definition video formats as specified in Section 6.2.

Legacy SD – A specific set of standard definition video formats as specified in Section 6.1.

Progressive Video – A specific set of formats as specified in Section 6.3.

Spatial Scalable Coding – A method of providing low and high spatial resolution versions of content through use of a base and an enhancement layer

Standard Frame Rate – A picture frequency lower than or equal to 60 pictures per second.

reference view – Stereoscopic 3D video component with the spatial resolution equal to or greater than the additional view.

reserved – Set aside for future use by a Standard.

Wide Color Gamut – A feature that allows representation of chrominance levels with much broader range than is possible with BT.709.

3.5 Extensibility

3.5.1 Backward-compatible Extensibility Mechanisms

This Standard includes no known backward-compatible extensibility mechanisms.

3.5.2 Non-backward-compatible Extensibility Mechanisms

Section 7 recommends signaling of the video characteristics that are to be specified in other parts of the ATSC 3.0 standard. Reserved fields in these signaling mechanisms will allow for extensibility.

3.5.3 Extensions with unknown compatibility

This Standard includes no known extensibility mechanisms of unknown compatibility.

3.5.4 Descriptor Processing Considerations

This Standard includes no known descriptor processing considerations.

4. SYSTEM CAPABILITIES (INFORMATIVE)

The HEVC video system for ATSC 3.0 supports Progressive Video resolutions as high as 3840 pixels by 2160 lines and picture rates as high as 120 fps. The system also supports Legacy SD and Interlace HD Video. Section 6.2 specifies video formats.

The system supports AFD and Bar Data (Sections 5 and 6.4) which support video of various aspect ratios.

For Progressive Video formats, the system supports various additional features as specified in Section 6.3. These are as follows:

- The system supports Spatial Scalable Coding (Section 6.3.1), which allows emission of a base layer with one resolution and a separate emission of an enhancement layer that, together with the base layer, provides a higher resolution result.
- The system supports various transfer characteristics (Section 6.3.2), including Standard Dynamic Range and various High Dynamic Range options. In addition, the system supports both Standard Color Gamut and Wide Color Gamut video.
- The system supports 3D video for both the Progressive Video and Interlaced HD Video formats as specified in Section 6.3.3.
- The system supports High Frame Rate video in ways that are compatible with SFR receivers and displays as specified in Section 6.3.4.

5. ACTIVE FORMAT DESCRIPTION

When the active image area of the emitted video signal does not fill the entire encoded video frame (e.g., when the video is letterboxed or pillarboxed), Active Format Description (AFD) and Bar Data information should be present in the original source video signal in accordance with SMPTE ST 2016-1 [6] and should be present in the emitted video signal. AFD information and Bar Data are used by receivers to optimize the display of images that do not fill the coded frame.

Bar Data values, when used, shall reflect the resolution, line and pixel counts of the original source pictures in the emitted video.

Bar Data values may be used for resolutions that are specified in this document but are not listed in Table 2 of Section 5.1 of SMPTE ST 2016-1 [5]. For convenience, Table 2 from SMPTE ST 2016-1 [6] (with anticipated UHDTV extensions) is reproduced below:

Table 5.1 SMPTE ST 2016-1 Table 2 – Video Format Information

Format	Applicable Production Standard	Pixels x Lines	Coded Pixels	Coded Lines		
				First Field	Second Field	Frame
480 Interlaced	SMPTE ST 125	720 x 480	0 – 719	23 – 262	286 – 525	–
480 Progressive	SMPTE ST 293	720 x 480	0 – 719	–	–	45 – 524
576 Interlaced	ITU-R BT.656	720 x 576	0 – 719	23 – 310	336 – 623	–
576 Progressive	ITU-R BT.1358	720 x 576	0 – 719	–	–	45 – 620
720 Progressive	SMPTE ST 296	1280 x 720	0 – 1279	–	–	26 – 745
1080 Interlaced	SMPTE ST 274	1920 x 1080	0 – 1919	21 – 560	584 – 1123	–
1080 Progressive	SMPTE ST 274	1920 x 1080	0 – 1919	–	–	42 – 1121
2160 Progressive	SMPTE ST 2036-1	3840 x 2160	0 – 3839	–	–	0 – 2159[2]
4320 Progressive	SMPTE ST 2036-1	7680 x 4320	0 – 7679	–	–	0 – 4319[2]

6. SOURCE CODING SPECIFICATION

This section establishes a specific subset of the HEVC video compression standard [2].

The constraints and specifications applicable to HEVC encoded ATSC 3.0 video bit streams are listed in Sections 6.1 through 6.4.

6.1 HEVC (ISO/IEC 23008-2)

The emitted coded video bitstream shall be compliant with ISO/IEC 23008-2 (HEVC) [2].

Each SPS shall have bit_depth_luma_minus8 and bit_depth_chroma_minus8 both equal to 0 or both equal to 2. This constrains the bit depth of luma and chroma to be 8 or 10 bits and the bit depth of luma and chroma to be identical.

When Spatial Scalable Coding is employed, the HEVC bitstream shall also conform to the constraints specified in 6.3.1.

6.2 Video Formats

The emitted HEVC encoded video streams shall be Legacy SD video, Interlaced HD video or Progressive Video as defined by 6.2.1, 6.2.2, and 6.2.3, respectively.

HDR, WCG, HFR, and Spatial Scalable Coding are not supported by Legacy SD or Interlaced HD video formats. These formats are included to maximize compatibility with the existing library of content that has an interlaced scanning structure and might also have an aspect ratio of 4:3. HDR, WCG, HFR, and Spatial Scalable Coding are supported by Progressive Video formats. 3D is supported by the Interlaced HD and Progressive Video formats.

6.2.1 Legacy SD Video

The ATSC 3.0 HEVC encoded Legacy SD video formats are specified in Table 6.1.

Emitted HEVC encoded video streams with the picture formats listed in Table 6.1 shall be encoded with the following constraints:

[2] These vertical line numbers are based on the active image area as specified in SMPTE ST 2036-1.

- The bitstream shall conform to HEVC Main 10 Profile, Main Tier, Level 3.1.
- The spatial resolution in both dimensions shall be evenly divisible by 8.
- In 60 Hz regions, the picture rates of 25 and 50 Hz shall not be used.
- In 50 Hz regions, the picture rates of 24/1.001, 24, 30/1.001, 30, 60/1.001, and 60 Hz shall not be used.
- Each SPS shall have vui_parameters_flag set equal to 1. This indicates that the vui_parameters() syntax structure as specified in HEVC [2] Annex E is present.
- Each SPS shall have colour_description_present_flag set equal to 1. This indicates that the colour_primaries, transfer_characteristics, and matrix_coeffs elements are present.
- Each SPS shall have colour_primaries present and set equal to 6. This constrains the color primaries to be ITU-R BT.601 [3].
- Each SPS shall have transfer_characteristics present and set equal to 6. This constrains the transfer characteristics to SDR as specified by ITU-R BT.601 [3].
- Each SPS shall have matrix_coeffs present and set equal to 6. This constrains the matrix coefficients to be ITU-R BT.601 [3].
- Each SPS shall have video_full_range_flag present and set equal to 0. This constrains the sample values to be of narrow range. Note that "narrow range" is sometimes referred to as "limited range."
- Spatial Scalable Coding (Section 6.3.1) shall not be used.
- HDR capabilities (Section 6.3.2) shall not be used.
- 3D capabilities (Section 6.3.3) shall not be used.
- High Frame Rate temporal sub-layering capabilities (Section 6.3.4) shall not be used.

Note that HEVC Main 10 Profile constrains the color subsampling to 4:2:0.

Table 6.1 Legacy SD Video Formats

Vertical Size (per field)	Vertical Size (per frame)	Horizontal Size	Pixel aspect ratio	Display Aspect Ratio	Allowed Picture Rates	Progressive/ interlaced
240	480	720[1]	10:11, 40:33	4:3, 16:9	5,6,8	I
N/A	480	720[1]	10:11, 40:33	4:3, 16:9	1,2,3,4,7	P
240	480	704	10:11, 40:33	4:3, 16:9	5,6,8	I
N/A	480	704	10:11, 40:33	4:3, 16:9	1,2,3,4,7	P
240	480	640	1:1, 4:3	4:3, 16:9	5,6,8	I
N/A	480	640	1:1, 4:3	4:3, 16:9	1,2,3,4,7	P

Legend:
picture rate: 1 = 23.976 Hz, 2 = 24 Hz, 3 = 29.97 Hz, 4 = 30 Hz, 5 = 59.94 Hz, 6 = 60 Hz, 7 = 25 Hz, 8 = 50 Hz, 9 = 100 Hz, 10 = 120/1.001 Hz, 11 = 120 Hz.
Footnotes:
1 Note that for 720x480 resolution formats, the active 4:3 or 16:9 picture falls within the center 704 pixels. The additional pixels account for the transitions created by analog blanking.

6.2.2 Interlaced HD Video

The ATSC 3.0 HEVC encoded Interlaced HD video formats are specified in Table 6.2.

Emitted HEVC encoded video streams with the picture formats listed in Table 6.2 shall be encoded with the following constraints:

- The bitstream shall conform to HEVC Main 10 Profile, Main Tier, Level 4.1.

- The spatial resolution in both dimensions shall be evenly divisible by 8.
- In 60 Hz regions, the picture rates of 25 and 50 Hz shall not be used.
- In 50 Hz regions, the picture rates of 24/1.001, 24, 30/1.001, 30, 60/1.001, and 60 Hz shall not be used.
- Each SPS shall have vui_parameters_flag set equal to 1. This indicates that the vui_parameters() syntax structure as specified in HEVC [2] Annex E is present.
- Each SPS shall have colour_description_present_flag set equal to 1. This indicates that the colour_primaries, transfer_characteristics, and matrix_coeffs elements are present.
- Each SPS shall have colour_primaries present and set equal to 1. This constrains the color primaries to be ITU-R BT.709 [4].
- Each SPS shall have transfer_characteristics present and set equal to 1. This constrains the transfer characteristics to SDR as specified by ITU-R BT.709 [4].
- Each SPS shall have matrix_coeffs present and set equal to 1. This constrains the matrix coefficients to be ITU-R BT.709 [4].
- Each SPS shall have video_full_range_flag present and set equal to 0. This constrains the sample values to be of narrow range. Note that "narrow range" is sometimes referred to as "limited range."
- Spatial Scalable Coding (Section 6.3.1) shall not be used.
- HDR capabilities (Section 6.3.2) shall not be used.
- High Frame Rate temporal sub-layering capabilities (Section 6.3.4) shall not be used.

Note that HEVC Main 10 Profile constrains the color subsampling to 4:2:0.

Table 6.2 Interlaced HD Video Formats

Vertical Size (per field)	Vertical Size (per frame)	Horizontal Size	Pixel aspect ratio	Display Aspect Ratio	Allowed Picture Rates	Progressive/ interlaced
540[1]	1080[1]	1920	1:1	16:9	5,6,8	I[2]
540[1]	1080[1]	1440	4:3	16:9	5,6,8	I[2]

Legend:
picture rate: 1 = 23.976 Hz, 2 = 24 Hz, 3 = 29.97 Hz, 4 = 30 Hz, 5 = 59.94 Hz, 6 = 60 Hz, 7 = 25 Hz, 8 = 50 Hz, 9 = 100 Hz, 10 = 120/1.001 Hz, 11 = 120 Hz.
aspect_ratio_idc: 1 = 1:1 [square samples]
Footnotes:
1 These formats shall be coded with a vertical size of 544 lines per field (1088 lines per frame) in order for the vertical resolution of each picture to be divisible by 8. The bottom 4 lines (8 lines per frame) shall be black.
2 Note that when telecine content is encoded, an inverse telecine process may be applied by the encoder, yielding a coded bitstream of 24 or 24/1.001 Hz progressive 1080x1920 or 1080x1440 pictures. These formats (24 or 24/1.001 Hz progressive) and other progressive formats are described in Section 6.2.3.

6.2.3 Progressive Video

The ATSC 3.0 HEVC encoded Progressive Video formats are specified in Section 6.2.3.1.

Emitted HEVC encoded video streams with the picture formats specified in Section 6.2.3.1 shall be encoded with the constraints specified in Section 6.2.3.2. Example Progressive Video formats are shown in Annex B.

6.2.3.1 Progressive Video Formats

- The spatial resolution shall be constrained to not more than 2160 lines and 3840 horizontal pixels.
- The spatial resolution in both dimensions shall be evenly divisible by 8.
- The picture rate in 60 Hz regions shall be one of the following in Hz: 24/1.001, 24, 30/1.001, 30, 60/1.001, 60, 120/1.001, 120.
- The picture rate in 50 Hz regions shall be one of the following in Hz: 25, 50, 100.
- The scan shall be progressive.
- The pixel aspect ratio shall be 1:1 (square pixels).

Coded representation of video with 1080 lines (e.g., 1920x1080) may be coded either as 1080 lines or as 1088 lines. When the video is coded as 1088 lines, the bottom 8 lines shall be black.

When 960x540 is emitted, it shall be coded as 960x544, and the bottom four lines shall be black.

6.2.3.2 Additional Constraints

HEVC encoded ATSC 3.0 Progressive Video shall comply with the following constraints:

- The bitstream shall conform to HEVC Main 10 Profile or HEVC Scalable Main 10 Profile, Main Tier, Level 5.2. Note that when a bitstream is indicated to conform to a level that is lower than Level 5.2, it is also considered as conforming to Level 5.2.
- Each SPS shall have vui_parameters_flag set equal to 1. This indicates that the vui_parameters() syntax structure as specified in HEVC [2] Annex E is present.
- Each SPS shall have chroma_loc_info_present_flag set equal to 1 and both chroma_sample_loc_type_top_field and chroma_sample_loc_type_bottom_field set equal to 2. This indicates that the top left chroma sample is aligned with the luma pixel in the top left corner.

Note that HEVC Main 10 Profile constrains the color subsampling to 4:2:0.

For transfer characteristics and associated component signal representation constraints, see Section 6.3.2.

6.3 Specific Constraints Regarding Additional Features

In addition to single layer, 2D, standard dynamic range, non-High Frame Rate video, it is possible to emit video with Spatial Scalable Coding, High Dynamic Range, Wide Color Gamut, High Frame Rate, and 3D features. These features are all available for progressive formats. In addition, the 3D feature is available for Interlaced HD formats.

6.3.1 Specific Constraints Regarding Spatial Scalable Coding

When HEVC Spatial Scalable Coding is employed, the bitstream shall comply with the following constraints:

6.3.1.1 General Constraints

- The bitstream shall contain exactly two layers, a base layer and an enhancement layer, and the value of vps_max_layers_minus1 of each VPS shall be set equal to 1.
- The base layer shall conform to HEVC Main 10 Profile and Main Tier.
- The enhancement layer shall conform to HEVC Scalable Main 10 Profile and Main Tier.

- Each of the base layer and the enhancement layer shall conform to one of the formats specified in Section 6.2.3.
- The spatial resolution of the enhancement layer shall be equal to x times that of the base layer both horizontally and vertically. The value of x shall be 1.5, 2, or 3.

6.3.1.2 Picture Rate Related Constraints

The following constraints result in a constant picture rate:

- The vps_vui_present_flag in each VPS shall be set equal to 1, pic_rate_present_vps_flag shall be set equal to 1, pic_rate_present_flag[i][j] shall be set equal to 1 and constant_pic_rate_idc[i][j] shall be set equal to 1 for all i, for all j.
- For the 'layer set'[3] to be carried in the video subsystem of this specification the list of allowed values for avg_pic_rate[i][j] shall be those values that indicate the picture rates defined in Section 6.2.3.1.
- The vui_parameters_present_flag in each SPS shall be set equal to 1, vui_timing_info_present_flag in each SPS shall be set equal to 1, vui_hrd_parameters_present_flag in each SPS shall be set equal to 1, and fixed_pic_rate_general_flag[i] shall be set equal to 1 or fixed_pic_rate_within_cvs_flag [i] shall be set equal to 1 for all values of i in the range 0 to maxNumSubLayersMinus1, inclusive.
- If vps_num_hrd_parameters is greater than 0, the following shall apply:
  - For each hrd_parameters() syntax structure in the VPS that applies to the layer set to be carried in the video subsystem of this specification: fixed_pic_rate_general_flag[i] shall be set equal to 1 or fixed_pic_rate_within_cvs_flag[i] shall be set equal to 1 for all values of i in the range 0 to maxNumSubLayersMinus1, inclusive.
  - For each value of i in the range 0 to maxNumSubLayersMinus1, inclusive, the value of the syntax element elemental_duration_in_tc_minus1[i] in the hrd_parameters() syntax structure applicable to the enhancement layer shall be equal to the value of the syntax element elemental_duration_in_tc_minus1[i] in the hrd_parameters() syntax structure applicable to the base layer.

6.3.1.3 Bit Depth, Dynamic Range, and Color Space Related Constraints

- The bit depth of luma, as specified by bit_depth_luma_minus8, shall be identical in the base layer and its corresponding enhancement layer.
- The bit depth of chroma, as specified by bit_depth_chroma_minus8, shall be identical in the base layer and its corresponding enhancement layer.
- The settings of Section 6.3.2 that are applied to the base layer shall be applied identically to the enhancement layer.
- The colour_mapping_enabled_flag shall be set to 0.

6.3.1.4 Random Access Related Constraints

- The random access picture period of the enhancement layer shall be equal to or greater than the random access picture period of the base layer. When an access unit[4] includes an IRAP picture with nuh_layer_id greater than 0, it shall also include an IRAP picture with nuh_layer_id equal to 0.

[3] The term 'layer set' is defined by the HEVC standard [2].

[4] The term 'access unit' is defined by the HEVC standard [2].

6.3.2 Specific Constraints Regarding Video Transfer Characteristics

The video transfer characteristics shall be signaled as one of the following: SDR, PQ, or HLG as specified in Sections 6.3.2.1, 6.3.2.2, and 6.3.2.3, respectively.

Note that switching between these transfer characteristics during a transmission could lead to a disturbance in downstream retransmission systems and consumer devices.

Conversions among different Transfer Characteristics (HDR/HDR, HDR/SDR) may be necessary for broadcast emission and/or MVPD retransmission to allow the Transfer function and color space to remain constant within or among services or for other purposes; some of these conversions will need to be made in real-time. Report ITU-R BT.2390 [29] contains information on conversions between HDR Transfer Characteristics and associated constraints and compromises.

6.3.2.1 SDR Transfer Characteristics

For video with SDR transfer characteristics, the following constraints apply:

- Each SPS shall have colour_description_present_flag set equal to 1. This indicates that the colour_primaries, transfer_characteristics, and matrix_coeffs elements are present.
- Each SPS shall have colour_primaries present and set equal to either 1 or 9. This constrains the color primaries to be ITU-R BT.709 [4] or ITU-R BT.2020 [5], respectively.
- Each SPS shall have transfer_characteristics present and set equal to 1. This constrains the transfer characteristics to SDR as specified by ITU-R BT.709 [4].
- Each SPS shall have matrix_coeffs present and set equal to either 1 or 9 and the value shall be identical to the value of colour_primaries. This constrains the matrix coefficients to be ITU-R BT.709 [4] or ITU-R BT.2020 [5] with non-constant luminance, respectively.
- Each SPS shall have video_full_range_flag present and set equal to 0. This constrains the sample values to be of narrow range. Note that "narrow range" is sometimes referred to as "limited range."
- When Spatial Scalable Coding is used, the constraints of Section 6.3.1 shall apply.

Video with SDR transfer characteristics may contain SL-HDR1 Metadata as documented in Section 6.3.2.1.1.

6.3.2.1.1 SL-HDR1 Metadata

SL-HDR1 Metadata messages are an aggregation of parameters, syntactically and semantically specified in [11], Section 6, which may be used in the decoding process to reconstruct HDR video from an SDR encoded video stream.

An HEVC or SHVC video stream may contain SL-HDR1 Metadata in order to provide both an SDR picture and an HDR picture from the same video stream. When SL-HDR1 Metadata messages are present, they allow reconstructing the HDR video from the received SDR video stream. The reconstructed HDR video can be represented as linear light or using any of the available HDR transfer functions listed in these specifications.

Usage of SL-HDR1 Metadata is as follows. At the emission encoder, HDR video is decomposed into SDR video ("derived SDR") and associated SL-HDR1 Metadata. The SDR video and accompanying SL-HDR1 Metadata are encoded in an HEVC stream.

When present, SL-HDR1 Metadata parameters are encapsulated into an SEI message, named SL-HDR Information SEI message, specified in [11], Section A.2.2. Mapping SL-HDR Information SEI message syntax elements to SL-HDR1 Metadata is documented in [11], Section A.2.3.

The following references document the non-normative SL-HDR1 decomposition process:

- Annex C of [11] provides HDR-to-SDR decomposition principles that may be used to generate the derived SDR video prior to encoding.
- Section D.3 of [11] provides the forward gamut mapping process that may be used to map the color gamut of the original HDR video (e.g., Rec. BT.2020) onto the target color gamut of the derived SDR video (e.g., Rec. BT.709). The forward gamut mapping process is not used when the HDR video gamut and the derived SDR video gamut are the same.

The following references document the normative SL-HDR1 reconstruction process for devices that support SL-HDR1:

- The SDR-to-HDR reconstruction process is specified in Section 7 of [11].
- When Gamut Mapping is employed at the encoder and signaled in the SL-HDR Information SEI message, the inverse gamut mapping process to map the color gamut of the derived SDR video (e.g., Rec. BT.709) back to the color gamut of the original HDR video (e.g., Rec. BT.2020) is specified in Section D.4 of [11].

If an SL-HDR Information SEI message is present, the following restrictions shall apply:

- It shall be transmitted at least with every IRAP.
- nal_unit_type shall be set to PREFIX_SEI_NUT.

  Note: This indicates that the SL-HDR Information SEI message is a prefix SEI NAL unit.

- sl_hdr_payload_mode shall be set to 0.

  Note: This constrains the payload carriage mode to the parameter-based mode.

- src_mdcv_info_present_flag value shall be set to 1.

  Note: This indicates that mastering display color volume primaries, white point and minimum and maximum luminance are present within the SL-HDR Information SEI message.

- When present, gamut_mapping_mode value shall be set to 1, 2 or 3.

  Note: This constrains the gamut mapping mode to explicit parameters, preset #1 (Rec. BT.709 to P3D65), and preset #2 (Rec. BT.709 to Rec. BT.2020), respectively.

6.3.2.1.2 SL-HDR1 Usage with SHVC

SL-HDR Information SEI messages may be used in combination with an SHVC two-layer spatial scalable stream as defined in Section 6.3.1. When this combination is employed, SL-HDR Information SEI messages may be present in the base layer, or present in the enhancement layer. If SL-HDR Information SEI messages are simultaneously present in both video layers, the SEI message in the enhancement layer is to be ignored; the base layer SEI message takes precedence.

When the SL-HDR Information SEI message is present in the base layer, the base and enhancement layers may both be decoded and reconstructed to HDR video by the SL-HDR1 reconstruction process as described above. When the SL-HDR Information SEI message is present in the enhancement layer, only the enhancement layer may be decoded and reconstructed to HDR video. The latter enables transmission of a video stream where the base layer is always decoded as SDR video while the spatial enhancement layer may be decoded as HDR video.

6.3.2.2 PQ Transfer Characteristics

For HDR video with the PQ transfer characteristics, the following constraints apply:

- Each SPS shall have colour_description_present_flag set equal to 1. This indicates that the colour_primaries, transfer_characteristics, and matrix_coeffs elements are present.
- Each SPS shall have colour_primaries present and set equal to 9. This constrains the color primaries to ITU-R BT.2100 [28].
- Each SPS shall have transfer_characteristics present and set equal to 16. This constrains the transfer characteristics to Table 4 (PQ System Reference Non-Linear Transfer Functions) of ITU-R BT.2100 [28].
- Each SPS shall have matrix_coeffs present and set equal to 9 or 14. This constrains the matrix coefficients to non-constant luminance $Y'C_BC_R$ or constant intensity IC_TC_P, respectively, as defined in ITU-R BT.2100 [28]. Note that switching between these modes during a transmission could lead to a disturbance in downstream retransmission systems and consumer devices.
- Each SPS shall have bit_depth_luma_minus8 and bit_depth_chroma_minus8 both set equal to 2. This constrains the bit depth to be 10 bits.
- Each SPS shall have video_signal_type_present_flag set equal to 1 and video_full_range_flag set equal to 0 or 1. This allows the sample values to be of narrow range or full range, respectively. For more information regarding digital representations, see SMPTE ST 2084 [9], Annex A. Note that "narrow range" is sometimes referred to as "limited range." Also note that switching between these modes during a transmission could lead to a disturbance in downstream retransmission systems and consumer devices.
- The bitstream may contain SEI messages with payloadType value equal to 137. This allows for the optional transmission of the Mastering Display Color Volume SEI message. It is not mandatory to transmit the Mastering Display Color Volume SEI message.
  - If the bitstream contains such messages, they shall conform to the construction described in HEVC [2].
- The bitstream may contain SEI messages with payloadType value equal to 144. This allows for the optional transmission of the Content Light Level Information SEI message. It is not mandatory to transmit the Content Light Level Information SEI message.
  - If the bitstream contains such messages, they shall conform to the construction described in HEVC [2].
- When Spatial Scalable Coding is used, the constraints of Section 6.3.1 shall apply.
- The bitstream may contain SEI messages with payloadType value equal to 4. This allows for the optional transmission of the ST 2094-10 metadata message described in Section 6.3.2.2.1 and the ST 2094-40 metadata message described in Section 6.3.2.2.2.

For more information regarding Mastering Display Color Volume metadata see SMPTE ST 2086 [26]. For more information regarding Content Light Level Information metadata (MaxFALL and MaxCLL) see CTA-861-G [27] Annex P.

6.3.2.2.1 Encoding and Transport of SMPTE ST 2094-10 Metadata Message

The HEVC video bitstream may contain the 2094-10 metadata message in order to provide dynamic information about the HDR video signal. When a 2094-10 metadata message is present, this information can be employed by the display to adapt the delivered HDR imagery to the

capability of the display device. Furthermore, this metadata message can be used to derive an SDR (ITU-R BT.709 [4]) picture by receiving devices such as an ATSC 3.0 receiver/converter. The information conveyed in the 2094-10 metadata message defined in Annex E provides carriage for metadata elements defined in ST 2094-1 [30] and ST 2094-10 [10].

2094-10 metadata, when present, shall be encoded and transported as User data registered by a Recommendation ITU-T T.35 Supplemental Enhancement Information (SEI) message per the ATSC1_data() structure defined in Table 14 of ANSI/SCTE 128-1 [7] and the assigned value for user_data_type_code is shown in Table 6.3.

Table 6.3 user_data_type_code

user_data_type_code	user_data_type_structure
0x09	ST2094-10_data()

The syntax and semantics for payload ST2094-10_data() shall be as specified in Annex E Section E.1. Where present, the corresponding NAL unit type shall be set equal to PREFIX_SEI_NUT.

If a 2094-10 metadata message is present, the following constraints shall apply:

- The 2094-10 metadata message shall be associated with every access unit of the bitstream. If this message is present, it shall only be present once per access unit.
- app_version shall be set equal to 0.
- Mastering Display Color Volume SEI messages (containing SMPTE ST 2086 [26] static metadata) shall be present in the bitstream as defined in Section 6.3.2.2.
- The number of extension blocks with ext_block_level equal to 1 shall be constrained to be equal to 1.
- The number of extension blocks with ext_block_level equal to 2 shall be constrained to be less than or equal to 16.
- The number of extension blocks with ext_block_level equal to 5 shall be constrained to be equal 0 or 1.

SCTE 215-1-1 2020b [31] defines an HEVC CMAF Media Profile that includes the carriage of SMPTE 2094-10:2016 dynamic metadata (see DM App #1).

6.3.2.2.2 Encoding and Transport of ST 2094-40 Metadata Messages

The HEVC video bitstream may contain ST 2094-40 metadata messages that provide for carriage of metadata elements defined in SMPTE ST 2094-1 [30] and SMPTE ST 2094-40 [13]. ST 2094-40 metadata messages, when present, provide descriptive statistical information and other information that can guide tone mapping in displays, frame by frame.

ST 2094-40 metadata, when present, shall be encoded and transported as User data registered by a Recommendation ITU-T T.35 Supplemental Enhancement Information (SEI) message per CTA-861-H [14] Annex S. The additional constraints specified in CTA-861.H [14] Annex S.4 shall apply. The syntax and semantics of the message are specified in CTA-861-H [14] Annex S.2 and Annex S.3, respectively.

When ST 2094-40 metadata messages are present, the following shall apply:

- The corresponding NAL unit type shall be set equal to PREFIX_SEI_NUT.
- One such message shall be associated with every access unit of the bitstream. If this message is present, it shall only be present once per access unit.

For information about the theory of operation for metadata based on SMPTE ST 2094-40, see Annex F.

SCTE 215-1-1 2020b [31] defines an HEVC CMAF Media Profile that includes the carriage of SMPTE 2094-40:2020 dynamic metadata (see DM App #4).

6.3.2.3 HLG Transfer Characteristics

For HDR video with the HLG transfer characteristics, the following constraints apply:

- Each SPS shall have colour_description_present_flag set equal to 1. This indicates that the colour_primaries, transfer_characteristics, and matrix_coeffs elements are present.
- Each SPS shall have colour_primaries present and set equal to 9. This constrains the color primaries to ITU-R BT.2100 [28].
- Each SPS shall have transfer_characteristics present and set equal to 18. This constrains the transfer characteristics to HLG as specified by Table 5 (Hybrid Log-Gamma (HLG) System Reference Non-Linear Transfer Functions) of ITU-R BT.2100 [28].
- Each SPS shall have matrix_coeffs present and set equal to 9. This constrains the matrix coefficients to non-constant luminance $Y'C_BC_R$, as defined in ITU-R BT.2100 [28].
- Each SPS shall have bit_depth_luma_minus8 and bit_depth_chroma_minus8 both set equal to 2. This constrains the bit depth to be 10 bits.
- Each SPS shall have video_signal_type_present_flag set equal to 1 and video_full_range_flag set equal to 0. This constrains the sample values to be of narrow range. For more information regarding signal representation, see ITU-R BT.2100 [28], Table 9. Note that "narrow range" is sometimes referred to as "limited range."
- When Spatial Scalable Coding is used, the constraints of Section 6.3.1 shall apply.

6.3.3 Specific Constraints Regarding 3D

The compression format for the reference view video and the additional view video shall be one of the formats described in Section 6.2.2 or 6.2.3. The compression formats of both views do not need to be identical. 3D video is not supported by the Legacy SD video format as defined by Section 6.2.1. If the reference view and the additional view have the same resolution, the reference view should be the left view.

6.3.3.1 Aspect Ratio

Both the reference and additional view videos shall have the same aspect ratio of the active area. If the aspect ratios of L/R pictures are not exactly identical, the smaller input pictures shall be letterboxed (or pillar boxed) before compression to make the both views have the same aspect ratio of the active area, and the existence of the bars embedded in the transmitted pictures shall be indicated by Active Format Description (AFD) and optionally Bar Data information as specified in Section 6.4.

6.3.3.2 Picture Rate

The picture rate of the additional view shall be that of the reference view divided by an integer of one or greater.

6.3.3.3 Multiview View Position SEI Message

When stereoscopic video is carried in SHVC[5], the multiview view position SEI message, the syntax of which is shown in Table 6.4, shall be used to indicate left and right view.

Table 6.4 Multiview View Position SEI Message

Syntax	Format
multiview_view_position(payloadSize) {	
num_views_minus1	ue(v)
for (i=0; i<=num_views_minus1; i++)	
view_position[i]	ue(v)
}	

The multiview view position SEI message as defined in [2] is identified by the payloadType value equal to 180. For the fixed and mobile hybrid 3D service, the view_position[i] in the SEI message indicates the order of the view with DependencyId equal to i among all the views from left to right for the purpose of display, with the order for the left-most view being equal to 0 and the value of the order increasing by 1 for next view from left to right. For example, when the base layer video is right view and the enhancement layer video is the left view, the view_position[0] will be 1 and the view_position[1] will be 0.

6.3.3.4 Multiview Scene Information SEI Message

The syntax of the multiview scene information SEI message shown in Table 6.5 may be used.

Table 6.5 Multiview Scene Information SEI Message

Syntax	Format
multiview_scene_info(payloadSize) {	
min_disparity	se(v)
max_disparity_range	ue(v)
}	

The payloadType of the multiview scene information SEI message shall be equal to 178 as defined in [2]. The multiview scene information SEI message could be used for processing the decoded views prior to rendering on a 3D display. The minimum disparity indicates the 3D location of the front-most object in the video. To render the receiver-generated graphics (e.g. OSD) in front of the decoded views on a 3D display, the minimum disparity in the multiview scene information SEI message should be included.

6.3.4 Specific Constraints Regarding High Frame Rate Temporal Sub-Layering

When an HEVC Main 10 Profile or HEVC Scalable Main 10 Profile bitstream has a constant picture rate (as indicated by the presence of elemental_duration_in_tc_minus1[ ]) equal to 120, 120/1.001, or 100 pictures per second (as specified by elemental_duration_in_tc_minus1[ ]), temporal sub-layering with two temporal sub-layers may be applied; otherwise, all the pictures shall have TemporalID equal to 0. To ensure that one sub-layer can be directly decoded by an HEVC Level 5.1 decoder, temporal sub-layering shall be applied to Level 5.2 HFR video streams.

[5] SHVC corresponds to the Scalable Main and Scalable Main 10 Profiles defined by the HEVC standard [2].

When temporal sub-layering with two temporal sub-layers is applied, the bitstream shall comply with the following constraints:

- The bitstream shall contain exactly two sub-layers, with TemporalId equal to 0 and 1, respectively, the value of sps_max_sub_layers_minus1 of each SPS shall be equal to 1, and the value of sps_max_sub_layers_minus1 of each VPS shall be equal to 1.
- The sub-layer representation with TemporalId equal to 0 shall have a constant picture rate (as indicated by the presence of elemental_duration_in_tc_minus1[0]), and the picture rate shall be exactly half of that of the entire bitstream (i.e., (elemental_duration_in_tc_minus1[0]+1) is equal to 2 * (elemental_duration_in_tc_minus1[1]+1)).
- In profile_tier_level( ) in each SPS, the value of sub_layer_level_present_flag[0] shall be equal to 1 only when the value of sub_layer_level_idc[0] is different from the value of general_level_idc. This constraint requires the signaling of the Level of the sub-layer representation with TemporalId equal to 0.
- The value of sub_layer_profile_present_flag[0] shall be equal to 1. This constraint requires the signaling of the profile related fields of the sub-layer representation with TemporalId equal to 0.

When temporal sub-layering with two temporal sub-layers is applied to the base layer and an enhancement layer[6] exists, the enhancement layer shall have the same picture rate as the picture rate of the base layer, and temporal sub-layering with two temporal sub-layers shall be applied to the enhancement layer with the same constraints as the base layer. The decoding order of each access unit in the sub-bitstream and each access unit in the subset shall be interleaved.

When temporal sub-layering with two temporal sub-layers is not applied to the base layer, all of the pictures of the enhancement layer shall have TemporalID equal to 0.

6.3.4.1 Specific Constraints Regarding Multiple Frame Rate Temporal Filtering

When High Frame Rate Temporal Sub-Layering is used, temporal filtering may be applied to the pictures of temporal sub-layer 0 prior to coding. This is intended to reduce or eliminate temporal aliasing (strobing) when temporal sub-layer 0 alone is displayed for the purposes of achieving Standard Frame Rate playback. In this case no temporal filtering shall be applied prior to coding the pictures of temporal sub-layer 1. Creation of a Multiple Frame Rate Temporal Filtering stream is described in Annex D.1.1.

Standard Frame Rate playback is achieved by rendering the pictures of temporal sub-layer 0 with no additional recovery process applied. High Frame Rate playback is achieved by rendering the pictures of temporal sub-layer 1 interspersed with the recovered, pre-filtered pictures of temporal sub-layer 0. Two parameters, temporal_filter_w1 and temporal_filter_w2 are used in the recovery process. temporal_filter_w1 shall range from 0 to 0.5, temporal_filter_w2 shall range from 1 to 0.5 and the sum of temporal_filter_w1 and temporal_filter_w2 shall be equal to 1 (i.e., temporal_filter_w2 is constrained to be equal to 1 minus temporal_filter_w1.) Annex D.1.2 describes recovery of the High Frame Rate stream consisting of recovered temporal sub-layer 0 and temporal sub-layer 1.

When temporal filtering is applied to the base layer and an enhancement layer[7] exists, temporal filtering shall be applied to the enhancement layer. Additionally, the value of temporal_filter_w1 shall be identical for both the base and enhancement layers, and the value of temporal_filter_w2 shall be identical for both the base and enhancement layers.

---

[6] Base and enhancement layers are used in Spatial Scalable Coding as described in 6.3.1.1.

[7] Base and enhancement layers are used in Spatial Scalable Coding as described in 6.3.1.1.

When temporal filtering is not applied to the base layer and an enhancement layer exists, temporal filtering shall not be applied to the enhancement layer.

For video streams in which all temporal sub-layers can be directly decoded by HEVC Level 5.1 decoders, temporal filtering shall not be applied.

Signaling is described in Section 7. For more information on Multiple Frame Rate Temporal Filtering streams and signaling, see Annex D.

6.4 Bit Stream Specifications in Addition to Video Coding

This section covers the specific data carried in the SEI and VUI sections of the video syntax of CVS (Coded Video Sequences) streams. The syntax used for the insertion of AFD and Bar Data, Carriage of Closed Captions, or Picture Timing Information in the SEI payload shall be as specified in the following sections..

6.4.1 Carriage of Active Format Description (AFD) and Bar Data

AFD and bar data, when present, shall be carried according to ANSI/SCTE 128-1 [7], Section 8.1 as further constrained and described in 6.4.2 of this document.

> Note: Additional background on AFD and bar data may be found in ETSI TS 101 154 V2.1.1 [20], the Digital Receiver Implementation Guidelines [22], and CTA-CEB16 [21].

6.4.2 AFD and Bar Data

AFD and bar data shall be as specified by ANSI/SCTE 128-1 [7], Section 8.2.

> Note (informative): The Display Aspect Ratio in Table 6.1, Table 6.2, and Table B.1.1 refers to the aspect ratio of the active image area before AFD and bar data are applied. Display devices can use AFD and bar data to crop and/or expand the picture that is displayed to the user. For example, a service including video encoded at a resolution of 3840x2160 may have AFD and bar data that indicates that the active image area is contained within a 3840x1620 rectangle.

> Note (informative): SMPTE 2016-1 does not include Ultra HD resolutions at this time and is in the process of being updated.

6.4.3 Optional Carriage of Closed Captions in the Video Stream

ATSC 3.0 specifies the broadcast carriage of closed captions as files that are formatted as "TTML Text and Image Profiles for Internet Media Subtitles and Captions (IMSC1)" [23] as specified by ATSC A/343 [25] (“Closed Captions and Subtitles Standard”). However, some broadcast signals are redistributed to viewers by networks that deliver closed captions as specified by CTA-708 [8] and carry the closed caption data embedded in the compressed video signal. In order to support interoperability with such networks, closed caption data may additionally be carried within the Supplemental Enhancement Information (SEI) of the HEVC video NAL unit. If this method is employed, carriage in this manner shall be as specified in Section 8 of ANSI/SCTE128-1 [7].

6.4.4 Carriage of Picture Timing Information with Coded Pictures

ATSC broadcasts may carry streams originated from mixed source types (e.g., interlace, progressive, frame repeated signals), in which case consistent availability of the Picture Timing Information SEI message at the decoder is necessary for the rendering engine to display decoded pictures correctly. In addition, in some regions these broadcast signals are likely to be redistributed to viewers by MVPD networks using the SCTE suite of standards, which requires that the Picture

Timing Information SEI be sent with each coded picture. The pic_struct, source_scan_type, duplicate_flag and other elements contained within the Picture Timing SEI message indicate whether a picture should be displayed as a frame, duplicate frames, or as one or more fields.

Picture Timing Information shall be carried within a prefix Picture Timing Information SEI message for each coded picture as specified in Section 7.1.7 of ANSI/SCTE 215-1 [12], and the frame_field_info_present_flag in the VUI shall be set to 1 as specified in Section 7.1.5 of ANSI/SCTE 215-1 [12]. In a single layer CVS stream, this is a single prefix SEI message associated with every coded picture. In a scalable stream, this is a single prefix SEI message associated with every coded picture in the base layer stream.

> Note (informative): Legacy deployments may exist which involve homogeneous source signal (e.g., progressive only) services, and in which no receiving entities in the ecosystem require the Picture Timing Information SEI message. Receiver implementations will ideally be able to decode streams that include Picture Timing Information SEI messages (legacy receivers may "gracefully ignore" the messages) and streams that do not include Picture Timing Information SEI messages.

7. SIGNALING OF VIDEO CHARACTERISTICS

Table 7.1 includes video characteristics that can be signaled.

Table 7.1 Video Characteristics

Item	Name	Description	Options	Reference
1	Codec	Indicates codec of the associated video component	HEVC	Section 6.1
2	Profile	Indicates profile of the associated video component	Main 10, Scalable; Main 10	Section 6.2
3	Tier	Indicates tier of the associated video component	Main[1]	Section 6.2
4	Level	Indicates level of the associated video component	3.1; 4.1; 5; 5.1; 5.2	Section 6.2
5	Color space container	Indicates the color space container of the associated video component	BT.709; BT.2020/BT.2100 (Refer to HEVC Specification [2])	Sections 6.2, 6.3.2
6	SCG compatibility	For a WCG (BT.2100) color space container, indicates that the content within that container is limited to SCG (BT.709) color gamut.	Non-compatible; Compatible	Section 6.3.2
7	dependencyComponentId	Indicates the component id of the video component that has to be decoded before decoding the associated video component	Component id of the base layer video	A/331 Annex B.9 [15]
8	Transfer Characteristics	Indicates the transfer characteristics of the video, including if it is SDR or HDR	SDR; PQ; HLG	Section 6.3.2
9	3D	Indicates whether the associated video components are part of 3D service	2D; 3D	Section 6.3.3
10	View Position	Indicates the view position of each video component of 3D service	Left; Right	Section 6.3.3
11	Scene disparity	Indicates the range of disparity within which the scene objects are located for 3D program	Minimum disparity: integer between -1024 and 1023, inclusive; Maximum disparity: integer between 0 and 2047, inclusive.	Section 6.3.3
12	Frame Rate	Indicates the frame rate of the video component	24/1.001; 24; 25; 30/1.001; 30; 50; 60/1.001; 60; 100; 120/1.001; 120	Section 6.2
13	Temporal Sub-Layering	Indicates whether 100, 120/1.001, and 120 Hz video is encoded using temporal sub-layering	Not using temporal sub-layering; Using temporal sub-layering	Section 6.3.4
14	Temporal Filtering	Indicates the parameters used in the temporal filtering recovery process.	temporal_filter_w1; temporal_filter_w2	Section 6.3.4.1

For information about signaling video characteristics, see ATSC A/331 [15].

Annex A: Video Inputs (Informative)

A.1 TELEVISION PRODUCTION STANDARDS

The television production standards and associated spatial resolutions shown in Table A.1.1 define video input formats that are intended to enhance interoperability in the overall use of the ATSC 3.0 Digital Television System.

Table A.1.1 Standard Video Production Formats

Video Standard	Active Lines	Active Samples/Line
SMPTE ST 2036-1 [15]	2160	3840
SMPTE ST 274 [17]	1080	1920
SMPTE ST 296 [18]	720	1280
SMPTE ST 125 [19]	483	720

Note that the emitted video format is not necessarily identical to the production format of the video input. For instance, in the case of pulldown, the input frame or field rate might be 60 frames per second, but the encoder might detect repeated frames and deliver a picture rate of 24 pictures per second for emission. Similarly, the video input signal might be decimated in order to provide an emitted video signal with a reduced resolution in order to provide a low bit rate result.

Annex B: Example Progressive Video Formats (Informative)

B.1 EXAMPLE PROGRESSIVE VIDEO FORMATS

Example Progressive Video formats are shown in Table B.1.1.

Table B.1.1 Example Progressive Video Formats

Vertical Size	Horizontal Size	aspect_ ratio_idc	Display Aspect Ratio	Picture Rates	Progressive/ interlaced
360	640	1	16:9	1,2,3,4,5,6,7,8	P
540	960[1]	1	16:9	1,2,3,4,5,6,7,8	P
720	1280	1	16:9	1,2,3,4,5,6,7,8,9,10,11	P
1080	1920	1	16:9	1,2,3,4,5,6,7,8,9,10,11	P
1440	2560	1	16:9	1,2,3,4,5,6,7,8,9,10,11	P
2160	3840	1	16:9	1,2,3,4,5,6,7,8,9,10,11	P

Legend:
picture rate: 1 = 23.976 Hz, 2 = 24 Hz, 3 = 29.97 Hz, 4 = 30 Hz, 5 = 59.94 Hz, 6 = 60 Hz, 7 = 25 Hz, 8 = 50 Hz, 9 = 100 Hz, 10 = 120/1.001 Hz, 11 = 120 Hz.
aspect_ratio_idc: 1 = 1:1 [square samples]
1 Note that when 960x540 is emitted, it is coded as 960x544, and the bottom four lines are black.

Annex C: Fixed/Mobile Hybrid UHD 3D Television System (Informative)

C.1 OVERVIEW

In case of UHD and HD simulcast broadcast delivery, the left view and the right view of a stereoscopic 3D video component can be a fixed UHD video and a mobile HD video, respectively. There is no dependency between two views as two views are coded independently. As shown in the example configuration of Figure C.1.1, each view can be delivered via separate PLPs (it can be differentiated according to ATSC 3.0 physical layer multiplexing schemes), and decoded independently. That is, a mobile receiver acquires HD right view video data from a PLP and offers a mobile HD service. A fixed receiver acquires UHD video data from a different PLP and provides a UHD service. Moreover, when a receiver can get data from two PLPs in ATSC 3.0 physical layer frames simultaneously, the receiver acquires an UHD left view and a HD right view video simultaneously and provides a 3D service by combining two views.

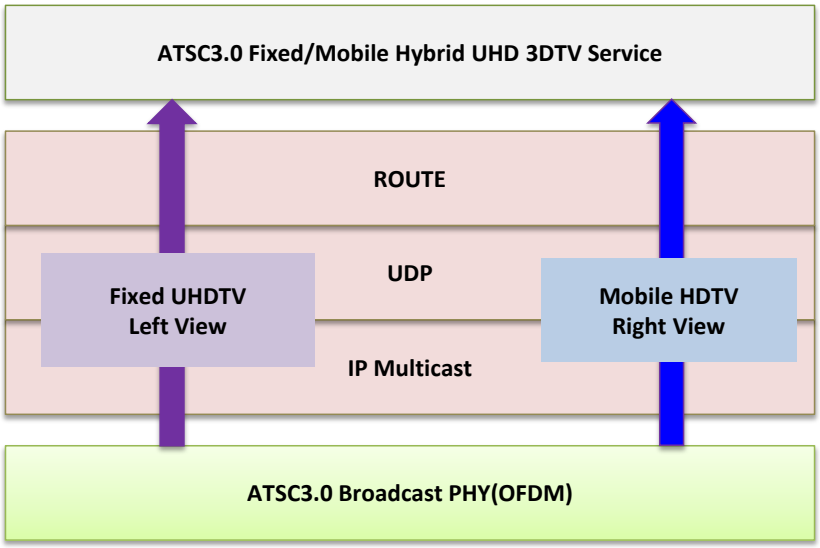

Figure C.1.1 Example configuration of a mobile HD right view and a fixed UHD left view.

By using layered coding (i.e., SHVC), the HD view can be coded in a base layer and the UHD view can be coded through both base and enhancement layers. Figure C.1.2 describes an example configuration for the fixed/mobile hybrid 3DTV service by using SHVC. Each layer can be delivered via same or separate PLPs. That is, the base layer can be delivered via a PLP for mobile service and the enhancement layer can be delivered via a different PLP. Thus, two PLPs, a more-robust one for a base layer and a less-robust higher-bandwidth one for an enhancement layer can be used for this SHVC-based approach. That is, a mobile receiver acquires HD right view video data from an ATSC 3.0 physical layer frame and offers a mobile HD service. Moreover, a fixed

receiver needs to acquire the UHD view by using both the base and the enhancement layers from two different PLPs. The receiver displays 3D service by combining two views.

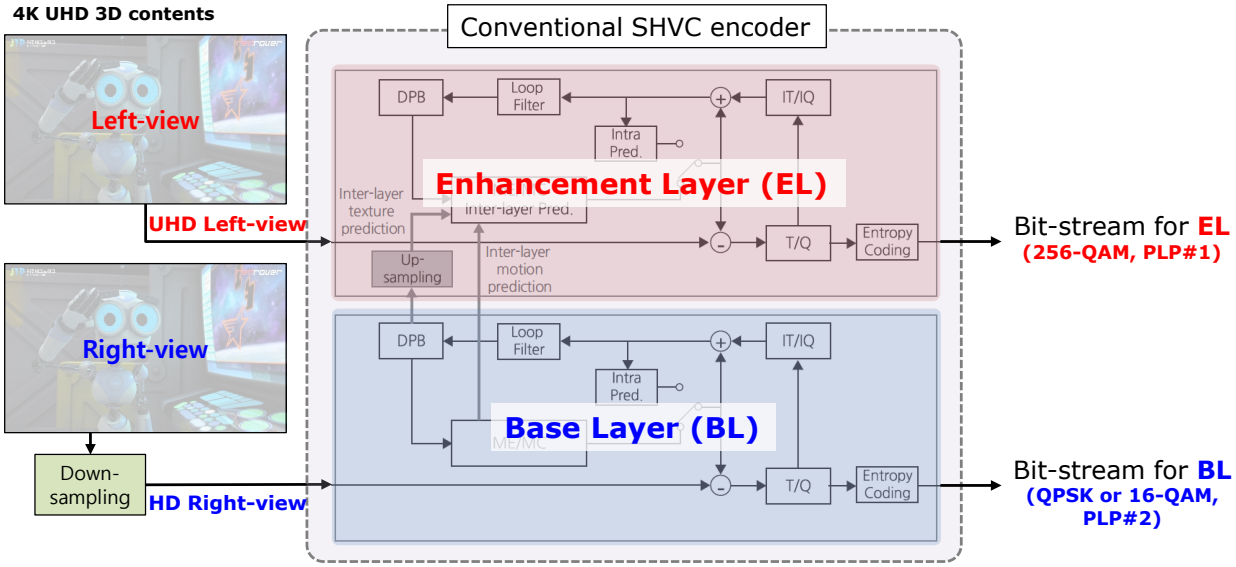

Figure C.1.2 Example configuration of a mobile HD right view and a fixed UHD left view using SHVC.

Annex D: Multiple Frame Rate Temporal Filtering (Informative)

D.1 MULTIPLE FRAME RATE TEMPORAL FILTERING

The Multiple Frame Rate Temporal Filtering tool allows efficient delivery of HFR and SFR video, each with somewhat independent effective shutter intervals. In the case that no temporal aliasing is desired, the ideal streams would be 120 fps video with a 1/120 s shutter interval along with 60 fps video with a 1/60 s shutter interval. (In 50 Hz regions, they would be 100 fps video with a 1/100 s shutter interval along with 50 fps video with a 1/50 s shutter interval.)

D.1.1 describes some possible methods of creating the two video streams as well as the use of the available parameters. D.1.2 describes methods of HFR video recovery from a Multiple Frame Rate Temporal Filtering stream.

D.1.1 Creation of a Multiple Frame Rate Temporal Filtering Stream

In the case that an HFR video stream is available, an SFR stream may be created by dropping every other picture; however, this may cause unwanted strobing (temporal aliasing.) This is especially true where a 1/120 (or 1/100) second shutter interval was initially used. The original video stream would have virtually no gaps in the motion blur, while dropping pictures would result in gaps in motion blur.

One can simulate a 1/60 (or 1/50) second shutter interval by averaging the current and previous pictures of the HFR video stream before decimating the stream with good results. However, if a shorter shutter interval was used (say 1/240s or 1/200s), blending the pictures evenly can cause double images. In this case, it can be desirable to use a weighted average of the pictures with more weight on the current picture than on the previous picture, balancing the negative effects of strobing (which can be strong with 60 fps at 1/250s) and the negative effects of a double image. By using this technique, the eye is drawn to the stronger image, rather than the secondary, ghost image, while the secondary image helps fill motion gaps.

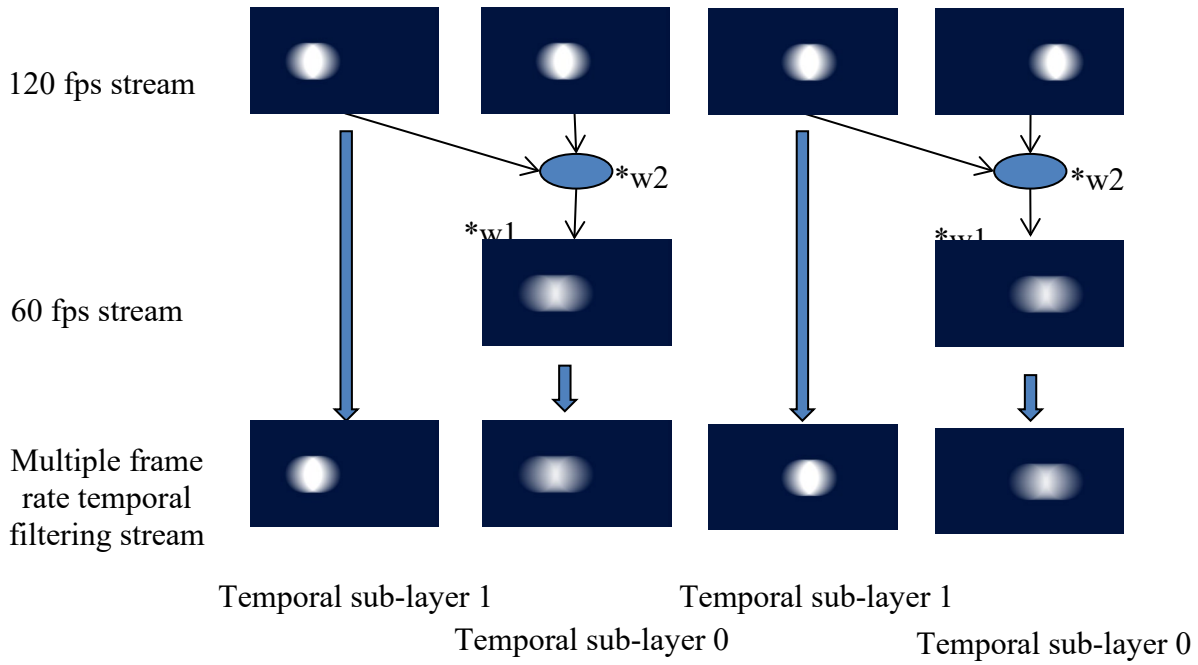

Figure D.1.1 Creation of a Multiple Frame Rate Temporal Filtering stream.

The creation of the SFR stream is not limited to the weighted averaging of pictures. A camera might provide two independent outputs or two cameras could be employed with an aligned, half-mirror rig. In the case of computer-generated graphics, the two streams could be rendered independently.

Regardless of how the two streams are generated, the pictures are interleaved in a single stream, using temporal sub-layering and two parameters are sent to indicate the intended method of HFR video recovery.

D.1.2 High Frame Rate Video Recovery

The SFR video is recovered by extracting the pictures of temporal sub-layer 0. No additional processing is necessary.

The HFR video is recovered by extracting the pictures of temporal sub-layer 1 and by processing the pictures in temporal sub-layer 0. The intended processing is indicated by parameters temporal_filter_w1 and temporal_filter_w2, which represent the pre-processing weights of the previous and current frames, respectively.

The available weighting factors were selected to allow efficient recovery, which can be implemented with shift, addition, and subtraction operations, rather than multiplication, divide or table lookup functions. A basic recovery scheme is shown as follows:

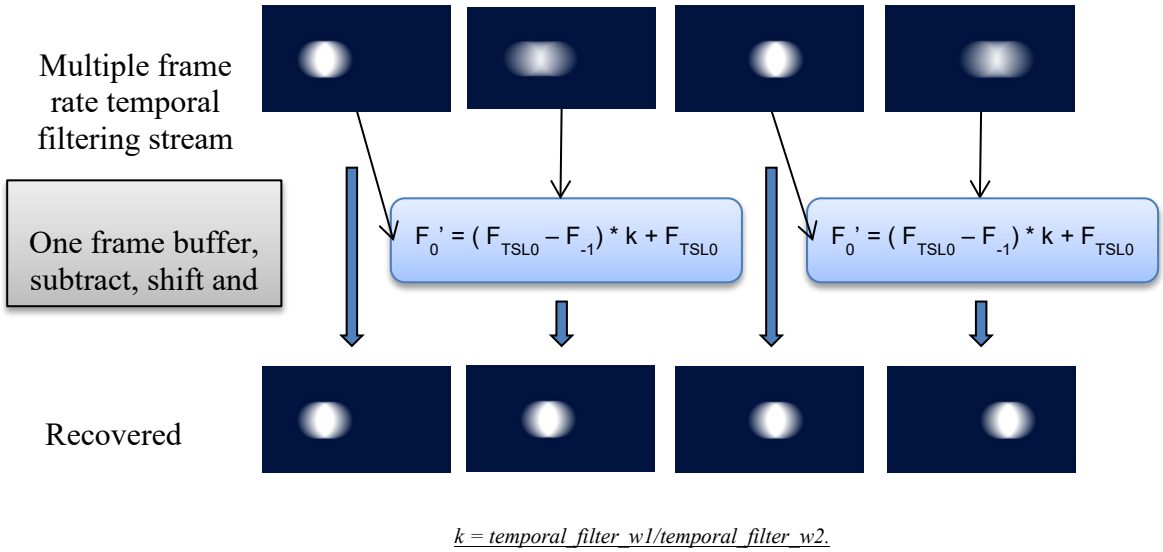

Figure D.1.2 Recovery of the HFR video stream.

In the above diagram, F_0' is the recovered picture and F_{-1} is the previous picture from temporal sub-layer 0. k = temporal_filter_w1/temporal_filter_w2.

In the case that the video will be displayed at frame rates that differ from that of the video stream, the displaying device can apply additional processing to achieve an optimized result.

Annex E: Metadata Based on SMPTE ST 2094-10_Data

E.1 METADATA BASED ON ST 2094-10_DATA (NORMATIVE)

This Section specifies the syntax and semantics of ST2094-10_data().

The syntax for ST2094-10_data() is shown in Table E.1.1, Table E.1.2, and Table E.1.3.

The parsing process of each syntax element by the descriptors f(n), i(n), ue(v), and u(n) is described in HEVC [2].

Note: The metadata elements are defined according to the SMPTE standards ST 2086 [26], ST 2094-1 [30], or ST 2094–10 [10]. Conversion between luminance values and 12-bit PQ values can be found in ST 2084 Annex A.

Table E.1.1 ST2094-10_data()

ST2094-10_data () {	**Descriptor**
app_identifier	ue(v)
app_version	ue(v)
metadata_refresh_flag	u(1)
if(metadata_refresh_flag) {	
num_ext_blocks	ue(v)
if(num_ext_blocks) {	
while(!byte_aligned())	
dm_alignment_zero_bit	f(1)
for(i = 0; i < num_ext_blocks; i ++) {	
ext_dm_data_block(i)	
}	
}	
}	
while(!byte_aligned())	
dm_alignment_zero_bit	
}	

Table E.1.2 ext_dm_data_block()

ext_dm_data_block(i) {	**Descriptor**
ext_block_length[i]	ue(v)
ext_block_level [i]	u(8)
ext_dm_data_block_payload(ext_block_length[i], ext_block_level [i])	
}	

Table E.1.3 ext_dm_data_block_payload()

ext_dm_data_block_payload(ext_block_length, ext_block_level) {	Descriptor
ext_block_len_bits = 8 * ext_block_length	
ext_block_use_bits = 0	
if(ext_block_level == 1) {	
min_PQ	u(12)
max_PQ	u(12)
avg_PQ	u(12)
ext_block_use_bits += 36	
}	
if(ext_block_level == 2) {	
target_max_PQ	u(12)
trim_slope	u(12)
trim_offset	u(12)
trim_power	u(12)
trim_chroma_weight	u(12)
trim_saturation_gain	u(12)
ms_weight	i(13)
ext_block_use_bits += 85	
}	
if(ext_block_level == 5) {	
active_area_left_offset	u(13)
active_area_right_offset	u(13)
active_area_top_offset	u(13)
active_area_bottom_offset	u(13)
ext_block_use_bits += 52	
}	
while(ext_block_use_bits++ < ext_block_len_bits)	
ext_dm_alignment_zero_bit	f(1)
}	

This section defines the semantics for ST2094-10_data().

For the purposes of the present section, the following mathematical functions apply:

$$\text{Abs(x)} = \begin{cases} x & ; \quad x \geq 0 \\ -x & ; \quad x < 0 \end{cases}$$

Floor(x) is the largest integer less than or equal to x.

$$\text{Sign(x)} = \begin{cases} 1 & ; \quad \text{x} > 0 \\ 0 & ; \quad x = 0 \\ -1 & ; \quad x < 0 \end{cases}$$

$$\text{Clip3(x, y, z)} = \begin{cases} x & ; \quad z < x \\ y & ; \quad z > y \\ z & ; \quad otherwise \end{cases}$$

$$\text{Round(x)} = \text{Sign(x)*Floor(Abs(x)+0.5)}$$

/ = Integer division with truncation of the result toward zero. For example, 7/4 and −7/−4 are truncated to 1 and −7/4 and 7/−4 are truncated to −1.

The precision of the information conveyed in this SEI message is intended to be adequate for purposes corresponding to the use of SMPTE ST 2094-10 [10].

app_identifier – Identifies an application and is set equal to 1 according to constraints of Section 5 of ST 2094-10 [10].

app_version – Specifies the application version in the application and is set equal to 0.

metadata_refresh_flag – When set equal to 1 cancels the persistence of any previous extended display mapping metadata in output order and indicates that extended display mapping metadata elements follow. The extended display mapping metadata persist from the coded picture to which the SEI message containing ST2094-10_data() is associated (inclusive) to the coded picture to which the next SEI message containing ST2094-10_data() and with **metadata_refresh_flag** set equal to 1 in output order is associated (exclusive) or (otherwise) to the last picture in the CVS (inclusive). When set equal to 0 this flag indicates that the extended display mapping metadata elements do not follow.

num_ext_blocks – Specifies the number of extended display mapping metadata blocks. The value shall be in the range of 1 to 254, inclusive.

dm_alignment_zero_bit – Shall be equal to 0.

ext_block_length[i] – Used to derive the size of the i-th extended display mapping metadata block payload in bytes. The value shall be in the range of 0 to 1023, inclusive.

ext_block_level[i] – Specifies the level of payload contained in the i-th extended display mapping metadata block. The value shall be in the range of 0 to 255, inclusive. The corresponding extended display mapping metadata block types are defined in Table E.1.4. Values of ext_block_level[i] that are ATSC reserved shall not be present in the bitstreams conforming to this version of ATSC specification. Blocks using ATSC reserved values shall be ignored.

When the value of ext_block_level[i] is set equal to 1, the value of ext_block_length[i] shall be set equal to 5.

When the value of ext_block_level[i] is set equal to 2, the value of ext_block_length[i] shall be set equal to 11.

When the value of ext_block_level[i] is set equal to 5, the value of ext_block_length[i] shall be set equal to 7.

Table E.1.4 Definition of Extended Display Mapping Metadata Block Type

ext_block_level	extended display mapping metadata block type
0	ATSC Reserved
1	Level 1 Metadata – Content Range
2	Level 2 Metadata – Trim Pass
3	ATSC Reserved
4	ATSC Reserved
5	Level 5 Metadata – Active Area
6…255	ATSC Reserved

When an extended display mapping metadata block with ext_block_level equal to 5 is present, the following constraints shall apply:

- An extended display mapping metadata block with ext_block_level equal to 5 shall be preceded by at least one extended display mapping metadata block with ext_block_level equal to 1 or 2.

- Between any two extended display mapping metadata blocks with ext_block_level equal to 5, there shall be at least one extended display mapping metadata block with ext_block_level equal to 1 or 2.
- No extended display mapping metadata block with ext_block_level equal to 1 or 2 shall be present after the last extended display mapping metadata block with ext_block_level equal to 5.
- The metadata of an extended display mapping metadata block with ext_block_level equal to 1 or 2 shall be applied to the active area specified by the first extended display mapping metadata block with ext_block_level equal to 5 following this block.
- When the active area defined by the current extended display mapping metadata block with ext_block_level equal to 5 overlaps with the active area defined by preceding extended display mapping metadata blocks with ext_block_level equal to 5, all metadata of the extended display mapping metadata blocks with ext_block_level equal to 1 or 2 associated with the current extended display mapping metadata block with ext_block_level equal to 5 shall be applied to the pixel values of the overlapping area.

min_PQ – Specifies the minimum luminance value of the current picture in 12-bit PQ encoding. The value shall be in the range of 0 to 4095, inclusive. Note that the 12-bit min_PQ value is calculated as follows:

$$\text{min_PQ} = \text{Clip3}(0, 4095, \text{Round}(\text{Min} * 4095))$$

where Min is MinimumPqencodedMaxrgb as defined in Section 6.1.3 of SMPTE ST 2094-10 [10].

max_PQ – Specifies the maximum luminance value of current picture in 12-bit PQ encoding. The value shall be in the range of 0 to 4095, inclusive. Note that the 12-bit max_PQ value is calculated as follows:

$$\text{max_PQ} = \text{Clip3}(0, 4095, \text{Round}(\text{Max} * 4095))$$

where Max is MaximumPqencodedMaxrgb as defined in Section 6.1.5 of SMPTE ST 2094-10 [10].

avg_PQ – Specifies the average PQ code value for the luminance of the picture in 12-bit PQ encoding. The value shall be in the range of 0 to 4095, inclusive. Note that the 12-bit avg_PQ value is calculated as follows:

$$\text{avg_PQ} = \text{Clip3}(0, 4095, \text{Round}(\text{Avg} * 4095))$$

where Avg is AveragePqencodedMaxrgb as defined in Section 6.1.4 of SMPTE ST 2094-10 [10].

target_max_PQ – Specifies the maximum luminance value of a target display in 12-bit PQ encoding. The value shall be in the range of 0 to 4095, inclusive. The target_max_PQ is the PQ encoded value of TargetedSystemDisplayMaximumLuminance as defined in Section 10.4 of SMPTE ST 2094-1 [30].

Note: This SEI message does not support signalling of TargetedSystemDisplayPrimaries, TargetedSystemDisplayWhitePointChromaticity, and TargetedSystemDisplayMinimumLuminance that are specified as mandatory in ST 2094-10 [10].

If there is more than one extended display mapping metadata block with ext_block_level equal to 2, those blocks shall have no duplicated target_max_PQ.

trim_slope – Specifies the slope metadata. The value shall be in the range of 0 to 4095, inclusive. If trim_slope is not present, it shall be inferred to be 2048. Note that the 12-bit slope value is calculated as follows:

$$\text{trim_slope} = \text{Clip3}(0, 4095, \text{Round}((S-0.5) * 4096))$$

where S is the ToneMappingGain as defined in Section 6.2.3 of SMPTE ST 2094-10 [10].

trim_offset – Specifies the offset metadata. The value shall be in the range of 0 to 4095, inclusive. If trim_offset is not present, it shall be inferred to be 2048. Note that the 12-bit offset value is calculated as follows:

$$\text{trim_offset} = \text{Clip3}(0, 4095, \text{Round}((O+0.5) * 4096))$$

where O is the ToneMappingOffset as defined in Section 6.2.2 of SMPTE ST 2094-10 [10].

trim_power – Specifies the power metadata. The value shall be in the range of 0 to 4095, inclusive. If trim_power is not present, it shall be inferred to be 2048. Note that the 12-bit power value is calculated as follows:

$$\text{trim_power} = \text{Clip3}(0, 4095, \text{Round}((P-0.5) * 4096))$$

where P is the ToneMappingGamma as defined in Section 6.2.4 of SMPTE ST 2094-10 [10].

trim_chroma_weight – Specifies the chroma weight metadata. The value shall be in the range of 0 to 4095, inclusive. If trim_chroma_weight is not present, it shall be inferred to be 2048. Note that the 12-bit chroma weight value is calculated as follows:

$$\text{trim_chroma_weight} = \text{Clip3}(0, 4095, \text{Round}((CW+0.5) * 4096))$$

where CW is the ChromaCompensationWeight as defined in Section 6.3.1 of SMPTE ST 2094-10[10].

trim_saturation_gain – Specifies the saturation gain metadata. The value shall be in the range of 0 to 4095, inclusive. If trim_saturation_gain is not present, it shall be inferred to be 2048. Note that the 12-bit saturation gain value is calculated as follows:

$$\text{trim_saturation_gain} = \text{Clip3}(0, 4095, \text{Round}((SG+0.5) * 4096))$$

where SG is the SaturationGain as defined in Section 6.3.2 of SMPTE ST 2094-10[10].

ms_weight – This field is reserved for future specification. This 13-bit signed integer shall be 0x1fff (-1).

active_area_left_offset, active_area_right_offset, active_area_top_offset, active_area_bottom_offset – Specify the selected pixels of the current picture, in terms of a rectangular region specified in

picture coordinates for the active area. The values shall be in the range of 0 to 8191, inclusive. See also ProcessingWindow of ST 2094-10 [10].

active_area_left_offset, active_area_right_offset, active_area_top_offset, active_area_bottom_offset represent the coordinates of UpperLeftCorner and LowerRightCorner constrained in Section 7.1 of ST 2094-10 [10] as follows:

$$\text{UpperLeftCorner} = (\text{active_area_left_offset}, \text{active_area_top_offset})$$
$$\text{LowerRightCorner} = (\text{XSize} - 1 - \text{active_area_right_offset}, \text{YSize} - 1 - \text{active_area_bottom_offset})$$

where Xsize is the horizontal resolution of the current picture and Ysize is the vertical resolution of the current picture.

ext_dm_alignment_zero_bit shall be equal to 0.

Annex F Metadata Based on SMPTE ST 2094-40

F.1 THEORY OF OPERATION (INFORMATIVE)

F.1.1 Metadata

The metadata based on SMPTE ST 2094-40 can be considered in three groups. The first is for identification purposes. The second describes a basis OOTF (optical-optical transfer function) for a producer-specified target peak luminance. This basis OOTF can be used to construct the guided OOTF, which defines the adaptation to the presentation display. The third describes statistical characteristics of the video signal. Other elements are constrained so that they are, or effectively are, unused.

Metadata elements associated with identification are not used for image processing. These elements include:

itu_t_t35_country_code
itu_t_t35_terminal_provider_code
itu_t_t35_terminal_provider_oriented_code
application_identifier
application_mode

Metadata elements associated with the basis OOTF include:

targeted_system_display_maximum_luminance
tone_mapping_flag[w]
knee_point_x[w]
knee_point_y[w]
num_bezier_curve_anchors[w]
bezier_curve_anchors[w][i]

Metadata elements associated with statistical characteristics include:

maxscl[w][i]
average_maxrgb[w]
num_distributions[w]
distribution_index[w][i]
distribution_values[w][i]
fraction_bright_pixels[w]

F.1.2 Basis OOTF

Prior to emission, the metadata associated with the basis OOTF is set. The basis OOTF is the transfer function for presentation at a single peak luminance point. This point is set within the peak luminance range of television sets that are available to viewers and is represented by targeted_system_display_maximum_luminance.

The basis OOTF curve depends on the content within the video images. In general, the basis OOTF compresses the dynamic range in signal ranges with lower information density and preserves contrast and details in signal ranges with higher information density. A well-constructed basis OOTF is reasonably faithful to the look of the original scene and has no discontinuities or sharp slope changes that might introduce visible artifacts into the images. Specific algorithms for creating a basis OOTF are outside of scope of this document.

The basis OOTF allows those responsible for content to view the baseline tone-mapped results for the target peak luminance level on a mastering monitor for quality control purposes. By providing this same basis OOTF to ST 2094-40-capable displays, these displays receive the same baseline as a common starting point for their individual tone mapping. OOTF curves based on SMPTE ST 2094-40 are composed by a linear part starting from (0, 0) and ending at a knee point (k_x, k_y) and a second part which is the Bezier curve with its anchor points starting with (k_x, k_y). These two parts are linked together as a smooth and continuous curve to avoid banding artifacts. Figure F.1.1 shows an example OOTF.

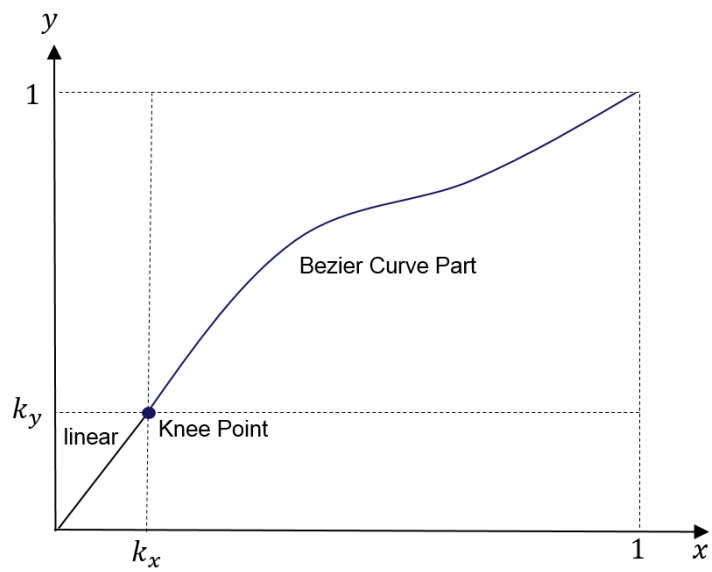

Figure F.1.1 Example OOTF.

See SMPTE ST 2094-40 [13] for more information regarding the basis OOTF.

F.1.3 Reference Method for Receiver-side Tone Mapping using ST 2094-40 Metadata

F.1.3.1 General

Figure F.1.2 shows a block diagram of a tone-mapping function based on ST 2094-40 metadata.

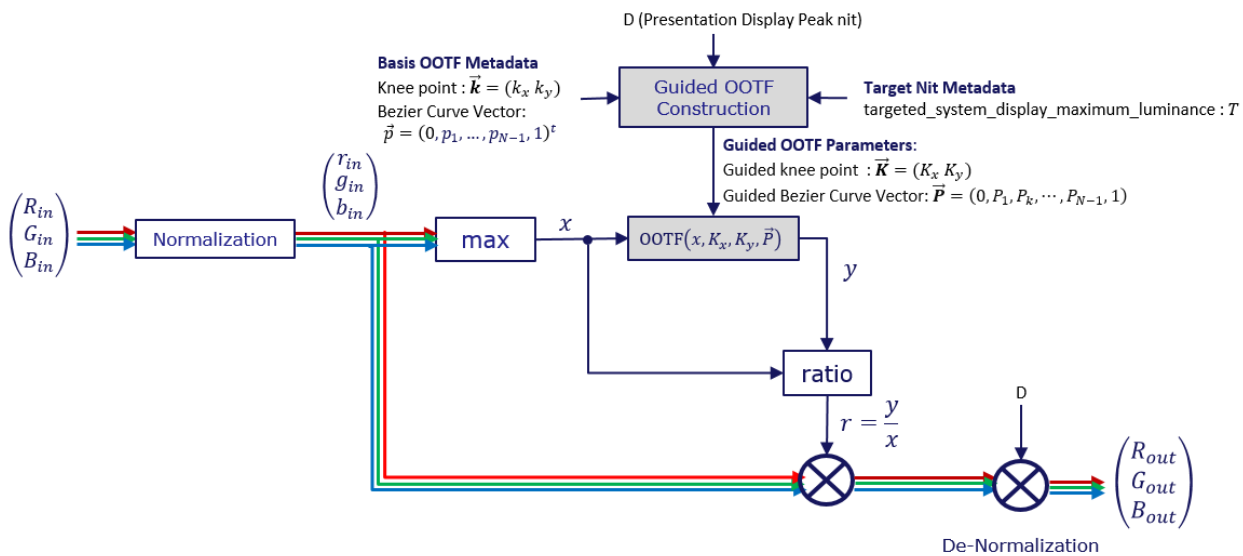

Figure F.1.2 Tone-mapping system of an ST 2094-40-based device.

Normalization is performed as follows:

The linear RGB signals $(R_{in}, G_{in}, B_{in})^t$ fed to the tone-mapping system represent absolute luminance, where t represents the vector transpose operation, and must be converted to a single channel input x before applying the guided OOTF, which processes a single input parameter.

The absolute luminance values $(R_{in}, G_{in}, B_{in})^t$ are normalized into the values between 0 and 1 by:

$$\begin{pmatrix} r_{in} \\ g_{in} \\ b_{in} \end{pmatrix} = \begin{pmatrix} \min(1, R_{in}/NORM\ ) \\ \min(1, G_{in}/NORM\ ) \\ \min(1, B_{in}/NORM\ ) \end{pmatrix} \tag{1}$$

where $NORM$ is the normalization factor given by:

$$NORM = \max(D, H_M) \tag{2}$$

in which D is the peak luminance of the presentation display and H_M = **distribution_values[0][8].**

For each pixel, the maximum value of r_{in}, g_{in}, and b_{in} for that pixel is determined, as represented by x. The value of x is applied to the guided OOTF, producing the resultant value, y. The values of r_{in}, g_{in}, and b_{in} are each scaled by the ratio of y/x.

At the end of the process, the signal is de-normalized based on the peak luminance of the presentation display.

F.1.3.2 Guided OOTF Construction

F.1.3.2.1 General

The guided OOTF is based on the peak luminance of the presentation display and is derived from the basis OOTF. There are three cases:

1) The peak luminance of the presentation display (D) is equal to the target peak luminance (T), in which case the basis OOTF can be used directly;
2) The peak luminance of the presentation display (D) is greater than the target peak luminance (T), in which case the basis OOTF is effectively interpolated with a linear transfer function to create the guided OOTF;
3) The peak luminance of the presentation display (D) is less than the target peak luminance (T), in which case the basis OOTF is extrapolated to create the guided OOTF.

In each case, the goals are to create a transfer function appropriate for the presentation display, to be reasonably faithful to the nature of the guided OOTF, and to avoid introducing level or slope discontinuities that might introduce visible artifacts.

A reference method of guided OOTF generation is as follows. This method is shown to produce good results; however, it is possible that other implementations can improve the interpolation and extrapolation processes to produce results with improved detail and faithfulness to the intent of the basis OOTF.

In general, guided OOTF construction is composed of the following two parts with the inputs T (peak luminance of the target display that is obtained with the basis OOTF) and D (peak luminance of the presentation display):

- Guided Knee Point
- Guided Bezier Curve Anchors

They are described in the subsequent sections.

F.1.3.2.2 Guided Knee Point Construction

The construction of the guided knee point can be classified into two cases.

Case I: When $D \leq T$

The guided knee point, $\vec{K} = \begin{pmatrix} K_x \\ K_y \end{pmatrix}$, can be obtained by:

$$\vec{K} = (w, \quad 1 - w) \cdot \begin{pmatrix} \vec{k} \\ \vec{K}_0 \end{pmatrix} \tag{3}$$

where $\cdot$ represents the dot product of two vectors, $\vec{k} = \begin{pmatrix} k_x \\ k_y \end{pmatrix}$ is the knee point of the given basis OOTF, $\vec{K}_0$ is a pre-defined constant vector such as $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$, and w is the guided knee point mixing parameter as a function of D. There are various ways to design w; however, a linear method is simple and effective. The linear method is as shown in Figure F.1.3, where D_L is a pre-defined low luminance level, and $D_L \leq T$.

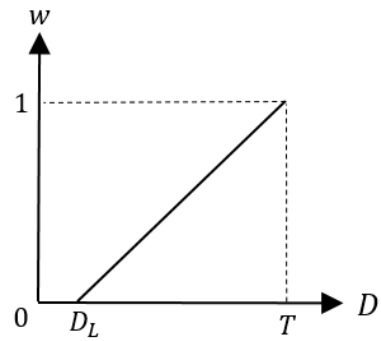

Figure F.1.3 Example of guided knee point mixing parameter function when $D \leq T$.

Figure F.1.4 shows a graphical illustration of how the guided knee point is constructed as a function of D as expressed in (3), where the dotted red arrow is the Guide Knee Point trajectory.

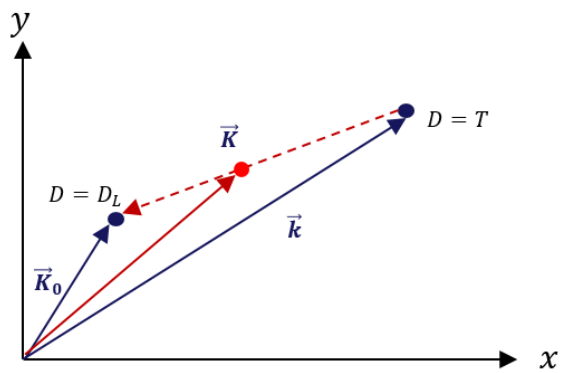

Figure F.1.4 The guided knee point when $D \leq T$.

Case II: When $T \leq D$

Similar to $D \leq T$, in the case that $T \leq D$, the guided knee point, $\vec{K} = \begin{pmatrix} K_x \\ K_y \end{pmatrix}$ can be obtained by:

$$\vec{K} = (w, \quad 1 - w) \cdot \begin{pmatrix} \vec{k} \\ \vec{K}_1 \end{pmatrix} \tag{4}$$

where $\vec{K}_1 = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$, and w is the guided knee point mixing parameter as a function of D which can be designed as shown in Figure F.1.5.

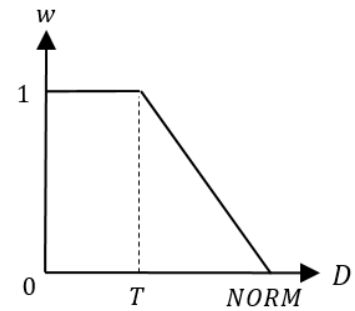

Figure F.1.5 Example of guided knee point mixing parameter function when $T \leq D$.

Figure F.1.6 shows a graphical illustration of how the guided knee point is constructed as a function of D as expressed in (4), where the dotted red arrow is the guide knee point trajectory.

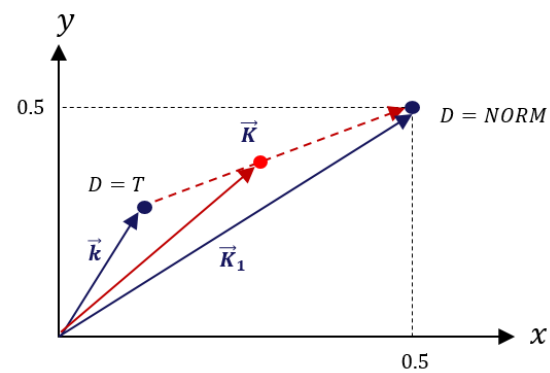

Figure F.1.6 The guided knee points when $T \leq D$.

Note that

$$\vec{K} = \vec{K}_1 = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} \tag{5}$$

when $NORM \leq D$.

F.1.3.2.3 Guided Bezier Curve Vector Construction

The following two properties of the Bezier curve are fundamental to understand the notion behind the guided Bezier curve vector construction.

Property 1: If:

$$\vec{P}_L = \begin{pmatrix} 0 \\ \frac{1}{N} \\ \vdots \\ \frac{N-1}{N} \\ 1 \end{pmatrix} \tag{6}$$

then the normalized explicit Bezier curve becomes an identity line. That is, $B_N(\vec{P}_L, t) = t$ and $\vec{P}_L$ is referred to as the identity Bezier curve vector.

Property 2: Linearly adding two Bezier curves in the same order is equivalent to linearly adding their Bezier curve vectors. That is, $B_N(\vec{\alpha}, t)$ and $B_N(\vec{\beta}, t)$ are two Bezier curves in N^{th} order characterized by the Bezier curve vectors $\vec{\alpha}$ and $\vec{\beta}$, respectively, such as:

$$\begin{cases} B_N(\vec{\alpha}, t) = \sum_{k=0}^{N} C_N^k \, t^k (1-t)^{N-k} \cdot \alpha_k \\ B_N(\vec{\beta}, t) = \sum_{k=0}^{N} C_N^k \, t^k (1-t)^{N-k} \cdot \beta_k \end{cases}, \tag{7}$$

then the following can be easily shown:

$$(a, b) \cdot \begin{pmatrix} B_N(\vec{\alpha}, t) \\ B_N(\vec{\beta}, t) \end{pmatrix} = B_N(\vec{\gamma}, t) \tag{8}$$

Where:

$$\vec{\gamma} = (a, b) \cdot \begin{pmatrix} \vec{\alpha} \\ \vec{\beta} \end{pmatrix} \tag{9}$$

Similar to the guided knee point construction, the guided Bezier curve vector construction from a given basis Bezier curve vector can also be classified into two cases. The basic idea behind guided Bezier curve construction is to interpolate a Bezier curve with the Bezier curve of the basis OOTF and a pre-determined boundary Bezier curve as a function of D.

Case I: When $D \leq T$

In the case of $D \leq T$, the guided Bezier curve, $B_N(\vec{P}, t)$, can be found as:

$$B_N(\vec{P}, t) = (u, 1-u) \cdot \begin{pmatrix} B_N(\vec{p}, t) \\ B_N(\vec{P}_0, t) \end{pmatrix} \tag{10}$$

where $\vec{p} = \vec{p} = (0,\ p_1,\ \ldots,\ p_{N-1},\ 1)^t$ is the Bezier curve vector of the basis OOTF, and $\vec{P}_0$ is a pre-defined Bezier curve vector such as, but not limited to, $(1,1,\cdots,1)^t$, and u is the control parameter as a function of D which can be designed as shown in Figure F.1.7 but not limited to.

By Property 2, the guided Bezier curve vector, $\vec{P}$, of $B_N(\vec{P}, t)$ can be calculated as:

$$\vec{P} = (u, \quad 1-u) \cdot \begin{pmatrix} \vec{p} \\ \vec{P}_0 \end{pmatrix} \tag{11}$$

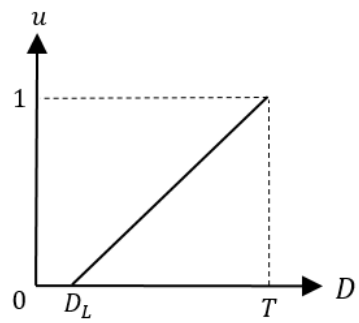

Figure F.1.7 Example of Bezier curve vector mixing coefficient function for $D \leq T$.

Case II: When $T \leq D$

In the case of $T \leq D$, the guided Bezier curve, $B_N(\vec{P}, t)$, can be found as:

$$B_N(\vec{P}, t) = (u, 1-u) \cdot \begin{pmatrix} B_N(\vec{p}, t) \\ B_N(\vec{P}_L, t) \end{pmatrix} \tag{12}$$

where $\vec{p} = \vec{p} = (0,\ p_1,\ \ldots,\ p_{N-1},\ 1)^t$ is the Bezier curve vector of the basis OOTF, $\vec{P}_L$ is the Identity Bezier curve introduced in (6), and u is the mixing parameter as a function of D which can be designed as shown in Figure F.1.8 but not limited to.

By Property 2, the guided Bezier curve vector, $\vec{P}$, of $B_N(\vec{P}, t)$ can be calculated as:

$$\vec{P} = (u, \quad 1-u) \cdot \begin{pmatrix} \vec{p} \\ \vec{P}_L \end{pmatrix} \tag{13}$$

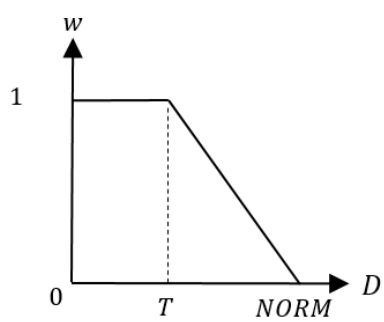

Figure F.1.8 The Bezier curve anchor control determining function for $T \leq D$.

Note that:

$$\vec{P} = \vec{P}_L \tag{14}$$

when $NORM \leq D$.

Note: By merging (5) and (14), the guided OOTF becomes the identity line as depicted in Figure F.1.9 when $NORM \leq D$, which is the case when the dynamic range of the presentation display is larger than that of the incoming HDR scene.

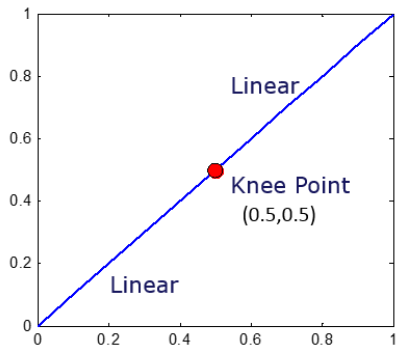

Figure F.1.9 The guided Bezier curve when $NORM \leq D$.

F.1.3.2.4 Slope Continuity Condition at the Knee Point

After the guided Knee point and Bezier curve vector are constructed, one last step is required to adjust the Bezier curve vector to ensure slope continuity at the guided knee point. Note that the linear part and the Bezier curve part of a guided OOTF curve are jointed by the knee point. Discontinuity between the two parts at the knee point may lead to banding artifacts in the tone-mapped signal. In any implementations of the ST 2094-40 tone-mapping system, the condition for slope continuity at the knee point must be satisfied. Namely, the tangent of the Bezier curve part at the knee point should be equal to the slope of the linear part, which can be expressed as:

$$\frac{d}{dx}\left(K_y + \left(1 - K_y\right) \cdot B_N\left(\vec{P}, \frac{x - K_x}{1 - K_x}\right)\right)\Bigg|_{x=K_x} = \frac{K_y}{K_x} \tag{15}$$

By solving (15), the condition for slope continuity at the knee point is given as:

$$P_1 = \frac{1}{N} \cdot \frac{K_y}{K_x} \cdot \frac{1 - K_x}{1 - K_y} \tag{16}$$

Figure F.1.10 shows some examples of various guided OOTF curves that satisfy the condition for slope continuity at the knee point. As can be seen in Figure F.1.10, the OOTF curves are continuous and will not lead to banding artifacts.

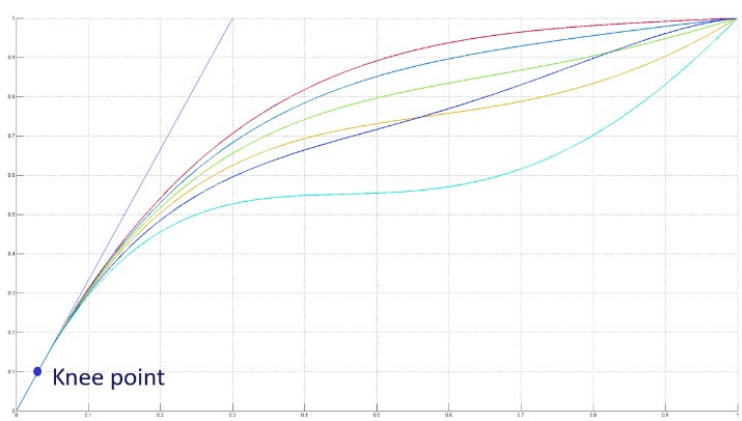

Figure F.1.10 Examples of the continuity of different guided OOTF curves at the knee point.

F.1.4 Statistical Characteristics

Statistical measurements of each video picture utilize straightforward mathematical processes that are performed before emission. See SMPTE ST 2094-40 [13] for details.

While it is possible for a receiver to make these same measurements on a frame-by-frame basis, it is more efficient to perform this process upstream. These video picture-based measurements cannot be completed until the full picture is available, potentially requiring frame buffers if calculated in the receiver. However, when measured before the emission, the calculations do not necessarily add latency. For instance, the measurement can be calculated in parallel with HEVC encoding, which necessarily adds latency of more than one picture of video, and the metadata messages are then inserted after encoding is completed. In addition, the receiver would not be able to make forward-looking, scene-based measurements.

Television sets, even with extremely high peak luminance capabilities, do not necessarily employ a neutral transfer function. The user might have selected a picture preset (Dynamic Mode) with high contrast based on personal preferences. An ambient light sensor might indicate a very bright environment, in which case low- and mid-tones might be lifted in order to ensure that the image is clearly visible. The manufacturer might also employ an adjustment to compensate for characteristics of the physical display device.

The statistical characteristic metadata elements can aid the television set in making the above adjustments. In general, a good implementation would avoid compressing the dynamic range in signal ranges where there is high information density and would instead compress the dynamic range in signal ranges with low information density. How this is achieved is left to the manufacturer and is not specified in this document.

– End of Document –